GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2002, 21:55:20 ; Search time 1732.53 Seconds
(without alignments)
3084.965 Million cell updates/sec

Title: US-09-785-276A-32
Perfect score: 396
Sequence: 1 ggcgaattggagctccccgc.....ttgtgaccagggcaagaggg 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	39.6	10.0	418	9	AI973196	AI973196 wr52h09.x
c 2	39	9.8	611	12	B75463	B75463 RPCI11-16G3
c 3	38.8	9.8	332	9	BE147008	BE147008 QV4-HT022
c 4	38.8	9.8	333	9	BE147013	BE147013 QV4-HT022
5	38.8	9.8	333	9	BE147015	BE147015 QV4-HT022
c 6	38.8	9.8	336	9	BE146800	BE146800 QV4-HT022
c 7	38.4	9.7	875	12	AZ534070	AZ534070 ENTBY52TF
c 8	38.4	9.7	929	12	AZ530962	AZ530962 ENTCH74TF
c 9	38.2	9.6	394	9	BB741171	BB741171 BB741171
c 10	37.8	9.5	495	10	BF392688	BF392688 UI-R-CA0-
c 11	37.6	9.5	340	9	BE146756	BE146756 QV4-HT022
c 12	37.6	9.5	341	9	BE147023	BE147023 QV4-HT022
c 13	37.6	9.5	342	9	BE146946	BE146946 QV4-HT022
c 14	37.6	9.5	702	9	AV722572	AV722572 AV722572
c 15	37.2	9.4	330	9	BE146803	BE146803 QV4-HT022
c 16	37.2	9.4	596	12	AZ974170	AZ974170 2M0248H08
17	36.8	9.3	342	9	AA723441	AA723441 zg69b05.s
c 18	36.8	9.3	710	12	BH607591	BH607591 BOGKT52TF
19	36.6	9.2	1085	12	CNS016YR	AL107373 Drosophil
20	36.4	9.2	521	9	AU146546	AU146546 AU146546
21	36.4	9.2	720	12	CNS015GI	AL105420 Drosophil
c 22	36.4	9.2	752	12	AZ614340	AZ614340 1M0443B11
23	36.4	9.2	942	9	AL519221	AL519221 AL519221
c 24	36.2	9.1	6146	12	AQ839831	AQ839831 260L13-C4
c 25	36	9.1	247	9	BB189866	BB189866 BB189866
c 26	36	9.1	337	9	BE147009	BE147009 QV4-HT022
c 27	35.8	9.0	207	12	AZ782914	AZ782914 2M0024K12
28	35.8	9.0	410	12	AZ115236	AZ115236 RPCI-23-1
c 29	35.6	9.0	450	9	AA526521	AA526521 ni60f10.s
30	35.6	9.0	604	9	AW559952	AW559952 EST315000
c 31	35.6	9.0	859	12	CNS004YY	AL055406 Drosophil
c 32	35.6	9.0	938	12	BH136715	BH136715 ENTOF89TR
33	35.4	8.9	298	9	BE147059	BE147059 QV4-HT022
c 34	35.4	8.9	404	10	BF297102	BF297102 047PbB05
c 35	35.4	8.9	544	12	AZ092469	AZ092469 RPCI-23-4
36	35.4	8.9	1086	12	CNS00YXR	AL096962 Drosophil
c 37	35.4	8.9	1204	12	CNS016E2	AL106628 Drosophil
c 38	35.2	8.9	235	9	AI480781	AI480781 ve47g08.x
c 39	35.2	8.9	308	9	AA240821	AA240821 mw24b07.r
c 40	35.2	8.9	515	12	AQ355100	AQ355100 CITBI-E1-
c 41	35.2	8.9	624	12	CNS01E33	AL140032 Anopheles
c 42	35.2	8.9	704	12	AQ355013	AQ355013 CITBI-E1-
43	35	8.8	452	10	H96455	H96455 yt98a03.r1
c 44	35	8.8	491	9	AV724923	AV724923 AV724923
c 45	35	8.8	590	12	AQ763239	AQ763239 HS_3222_B

ALIGNMENTS

RESULT 1
AI973196/c
LOCUS AI973196 418 bp mRNA linear EST 08-MAR-2000
DEFINITION wr52h09.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2491361 3', mRNA sequence.
ACCESSION AI973196
VERSION AI973196.1 GI:5770022
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 418)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 2291 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES Location/Qualifiers
source 1. .418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2491361"
/clone_lib="NCI_CGAP_Ut1"
/tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"
BASE COUNT 109 a 72 c 91 g 146 t
ORIGIN

Query Match 10.0%; Score 39.6; DB 9; Length 418;
Best Local Similarity 53.2%; Pred. No. 1.7;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 39 tctgaattacaatagatgtgctaagatggtatgtttcggtgtgtattctttggttgcaag 98
Db 159 TCCAAACTTCAAGGCATCTACTAAAATTATATTTATCTCTCTTTGATATACCAATCCCGC 100

Qy 99 caccaggcatgagttttggtgaatttaatcaagagagtaattcgttacaagaacttgggg 158
Db 99 AACTTTAAACTATTTATGCAGAAACTAATCTACATATTAAATTGTGCCAATAAATGTTGT 40

Qy 159 tttctgtgcaacccaagaaaagtctgaaaataaccaaa 196
Db 39 CTTCTCTGTTTCCATAAAAAAAAAAAAAAAAAAAAAAA 2

RESULT 2
B75463/c
LOCUS B75463 611 bp DNA linear GSS 08-APR-1999
DEFINITION RPCI11-16G3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-16G3, DNA sequence.
ACCESSION B75463
VERSION B75463.1 GI:2771150
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 611)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI11-16G3.TP
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .611
/organism="Homo sapiens"
/db_xref="GDB:7505906"
/db_xref="taxon:9606"
/clone="RPCI-11-16G3"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
BASE COUNT 164 a 139 c 126 g 182 t
ORIGIN

Query Match 9.8%; Score 39; DB 12; Length 611;
Best Local Similarity 51.4%; Pred. No. 2.7;
Matches 90; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 65 tggtatgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaattt 124
Db 541 TGGGATGGCAGACCATAGACTTCATGGTGAAAACTACCAAATATGACTTTAGCTAACTTA 482

Qy 125 aatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtctg 184
Db 481 AGTAAAAAAATTCATGTATTTGAAATATCTTGGTTCGCTTTAAGAATGAACAAAAGTTCT 422

Qy 185 aaaataaccaaatttagaaaatgtaaggaatcaggggttcaagactctagata 239
Db 421 AGAGAAACAAGTTTCAGAAATGAACAGTAGTCAAGGGTAGCTAGGTGGCTGGCTA 367

RESULT 3
BE147008/c
LOCUS BE147008 332 bp mRNA linear EST 21-JUN-2000
DEFINITION QV4-HT0222-011199-019-d12 HT0222 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE147008
VERSION BE147008.1 GI:8609732
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 332)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV4-HT0222-011199-019-d12&t3=1999-11-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 332.
FEATURES Location/Qualifiers
source 1. .332
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0222"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 99 a 74 c 93 g 66 t
ORIGIN

Query Match 9.8%; Score 38.8; DB 9; Length 332;
Best Local Similarity 51.8%; Pred. No. 2.8;
Matches 88; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

```
Qy    191 accaaatttagaaaatgtaaggaatcagggggtttcaagactctagatagcaagaagtgg 250
          | | |||||    | |   || |   ||||||||||  | | | | |||   ||  | |
Db    227 AGCTAATTTTTGTATTTGTAGTAGAGAGGGGGTTTCACCATTTTGGCTAGGCTGATCTTG 168

Qy    251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcca 310
          ||   |  | ||| || || || ||      |    ||      |    |||||| |  ||
Db    167 AACTCCTGGCCTCAAGCGATCCACCCGCCTTGGCCTTCCAAAGCTCTGGGTTTACAGGCA 108

Qy    311 agtgccattctccccaacctccaggagtgagagtttctttagtatgtctt 360
          | |||| | | |||| |||       | |    || | |  | ||| ||
Db    107 TGAGCCACTGTGCCCAGCCTGTTTTTTTTATTTGTTTTGTTTTTTGTTTT 58


RESULT  4
BE147013/c
LOCUS       BE147013                333 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  QV4-HT0222-011199-019-g05 HT0222 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE147013
VERSION     BE147013.1  GI:8609737
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 333)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV4-HT0222-011
            199-019-g05&t3=1999-11-01&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 6
            High quality sequence stop: 333.
FEATURES             Location/Qualifiers
     source          1. .333
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0222"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      100 a     75 c     94 g     64 t
ORIGIN


  Query Match             9.8%;  Score 38.8;  DB 9;  Length 333;
  Best Local Similarity   51.8%;  Pred. No. 2.8;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy    191 accaaatttagaaaatgtaaggaatcagggggtttcaagactctagatagcaagaagtgg 250
          | | |||||    | |   || |   ||||||||||  | | | | |||   ||  | |
Db    228 AGCTAATTTTTGTATTTGTAGTAGAGAGGGGGTTTCACCATTTTGGCTAGGCTGATCTTG 169

Qy    251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcca 310
          ||   |  | ||| || || || ||      |    ||      |    |||||| |  ||
Db    168 AACTCCTGGCCTCAAGCGATCCACCCGCCTTGGCCTTCCAAAGCTCTGGGTTTACAGGCA 109

Qy    311 agtgccattctccccaacctccaggagtgagagtttctttagtatgtctt 360
          | |||| | | |||| |||       | |    || | |  | ||| ||
Db    108 TGAGCCACTGTGCCCAGCCTGTTTTTTTTATTTGTTTTGTTTTTTGTTTT 59


RESULT  5
BE147015
LOCUS       BE147015                333 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  QV4-HT0222-011199-019-h06 HT0222 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE147015
VERSION     BE147015.1  GI:8609739
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 333)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV4-HT0222-011
            199-019-h06&t3=1999-11-01&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 333.
FEATURES             Location/Qualifiers
     source          1. .333
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0222"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       65 a     92 c     76 g     100 t
ORIGIN


  Query Match             9.8%;  Score 38.8;  DB 9;  Length 333;
  Best Local Similarity   51.8%;  Pred. No. 2.8;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy    191 accaaatttagaaaatgtaaggaatcagggggtttcaagactctagatagcaagaagtgg 250
          | | |||||    | |   || |   ||||||||||  | | | | |||   ||  | |
Db    125 AGCTAATTTTTGTATTTGTAGTAGAGAGGGGGTTTCACCATTTTGGCTAGGCTGATCTTG 184
```

```
Qy    251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcca 310
          ||   |  | ||| || || || ||      |    ||     |   |||||| |  ||
Db    185 AACTCCTGGCCTCAAGCGATCCACCCGCCTTGGCCTTCCAAAGCTCTGGGTTTACAGGCA 244

Qy    311 agtgccattctccccaacctccaggagtgagagtttctttagtatgtctt 360
           | |||| | | |||| |||      | |    || | |  | ||| ||
Db    245 TGAGCCACTGTGCCCAGCCTGTTTTTTTATTTGTTTTGTTTTTTGTATT 294


RESULT  6
BE146800/c
LOCUS       BE146800                336 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  QV4-HT0222-181099-013-b04 HT0222 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE146800
VERSION     BE146800.1  GI:8609524
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 336)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QV4-HT0222-181
            099-013-b04&t3=1999-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 9
            High quality sequence stop: 336.
FEATURES             Location/Qualifiers
     source          1. .336
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0222"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      100 a     77 c     96 g     63 t
ORIGIN


  Query Match             9.8%;  Score 38.8;  DB 9;  Length 336;
  Best Local Similarity   51.8%;  Pred. No. 2.8;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy    191 accaaatttagaaaatgtaaggaatcaggggttcaagactctagatagcaagaagtgg 250
          | | |||||    | |   || |   ||||||||||  | | | | |||   ||  | |
Db    231 AGCTAATTTTTGTATTTGTAGTAGAGAGGGGGTTTCACCATTTTGGCTAGGCTGATCTTG 172

Qy    251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcca 310
          ||   |  | ||| || || || ||      |    ||     |   |||||| |  ||
Db    171 AACTCCTGGCCTCAAGCGATCCACCCGCCTTGGCCTTCCAAAGCTCTGGGTTTACAGGCA 112

Qy    311 agtgccattctccccaacctccaggagtgagagtttctttagtatgtctt 360
           | |||| | | |||| |||      | |    || | |  | ||| ||
Db    111 TGAGCCACTGTGCCCAGCCTGTTTTTTTATTTGTTTTGTTTTTTGTTTT 62


RESULT  7
AZ534070/c
LOCUS       AZ534070                875 bp    DNA     linear   GSS 03-NOV-2000
DEFINITION  ENTBY52TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ534070
VERSION     AZ534070.1  GI:11090344
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 875)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 782.
FEATURES             Location/Qualifiers
     source          1. .875
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      316 a    165 c     96 g    298 t
ORIGIN


  Query Match             9.7%;  Score 38.4;  DB 12;  Length 875;
  Best Local Similarity   52.5%;  Pred. No. 4.1;
  Matches   84;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     78 tgtgtattctttggttgcaagcaccaggcatgagttttggtgaatttaatcaagagagta 137
          | |  |||| || |||| |    |   |  | ||| ||  | | ||| ||   ||
Db    396 TATTGATTCGTTTGTTGAAGATTCTTTGGTTAAGTATTTTTTATTTTTATATTGAATAGT 337

Qy    138 attcgttacaagaacttggggtttctgtgcaacccaagaaaagtctgaaaataaccaaat 197
           ||  |   || ||  |  ||||| | |  ||     || |  | ||  ||| |  ||||
Db    336 TTTTATATTAAAAAGATATGGTTTATTTCAAAGTTGTGAGATTTTTGTGAATGAAAAAAT 277

Qy    198 ttagaaaatgtaaggaatcaggggttcaagactctaga 237
```

```
          ||| | |   |||  ||| ||    |  | | | ||||||
Db    276 TTAAACACCTTAAAAAATGAGATTTTAACGAAATTCTAGA 237


RESULT   8
AZ530962/c
LOCUS       AZ530962                929 bp    DNA     linear   GSS 03-NOV-2000
DEFINITION  ENTCH74TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ530962
VERSION     AZ530962.1  GI:11085208
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 929)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 824.
FEATURES             Location/Qualifiers
     source          1. .929
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (-2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      328 a    172 c    106 g    323 t
ORIGIN


  Query Match             9.7%;  Score 38.4;  DB 12;  Length 929;
  Best Local Similarity   52.5%;  Pred. No. 4.2;
  Matches   84;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy      78 tgtgtattctttggttgcaagcaccaggcatgagttttggtgaatttaatcaagagagta 137
           | |  |||| || |||| |    |   |  | ||| ||  | | ||| ||   ||
Db     294 TATTGATTCGTTTGTTGAAGATTCTTTGGTTAAGTATTTTTTATTTTTATATTGAATAGT 235

Qy     138 attcgttacaagaacttggggtttctgtgcaacccaagaaaagtctgaaaataaccaaat 197
             ||  |   || ||  | ||||| | |  ||    || |  | ||  ||| |  ||||
Db     234 TTTTATATTAAAAAGATATGGTTTATTTCAAAGTTGTGAGATTTTTGTGAATGAAAAAAT 175

Qy     198 ttagaaaatgtaaggaatcaggggggtttcaagactctaga 237
           ||| | |   |||  ||| ||    |  | | | ||||||
Db     174 TTAAACACCTTAAAAAATGAGATTTTAACGAAATTCTAGA 135
```

```
RESULT   9
BB741171/c
LOCUS       BB741171                394 bp    mRNA    linear   EST 15-OCT-2001
DEFINITION  BB741171 RIKEN full-length enriched, 6 days neonate spleen Mus
            musculus cDNA clone F430213P21 3', mRNA sequence.
ACCESSION   BB741171
VERSION     BB741171.1  GI:16144176
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="F430213P21"
                     /clone_lib="RIKEN full-length enriched, 6 days neonate
                     spleen"
                     /tissue_type="spleen"
                     /dev_stage="6 days neonate"
BASE COUNT      107 a     84 c     87 g    116 t
ORIGIN


  Query Match             9.6%;  Score 38.2;  DB 9;  Length 394;
  Best Local Similarity   53.7%;  Pred. No. 4.2;
  Matches   79;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      94 gcaagcaccaggcatgagttttggtgaatttaatcaagagagtaattcgttacaagaact 153
           || ||| ||   |  ||   ||  ||  |  | ||| |  ||    ||  || | |||
```

```
Db    278 GCCAGCTCCTTACCAGACGGTTTCTGTTTAAATTCAGGCAAGATTGTCTCTAGATGAAAA 219

Qy    154 tggggtttctgtgcaacccaagaaaagtctgaaaataaccaaatttagaaaatgtaagga 213
          |  |||||  |    ||    |||| | ||   | || | |||  ||||| |||||| |
Db    218 TTTTGTTTCCATCTTCCCTTTAAAAATTATGTTCAAAATCTAATGAAGAAAGTGTAAGTA 159

Qy    214 atcaggggtttcaagactctagatag 240
          |||| |   |||  |||   | ||
Db    158 ATCATGTAAATTCTGTACTTAGCAGAG 132


RESULT 10
BF392688/c
LOCUS       BF392688                495 bp    mRNA    linear   EST 27-NOV-2000
DEFINITION  UI-R-CA0-bfm-e-11-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
            UI-R-CA0-bfm-e-11-0-UI 3', mRNA sequence.
ACCESSION   BF392688
VERSION     BF392688.1  GI:11377585
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized corpus-striatum library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 28-51, >ID_RN#SINE/ID
            112-144, >(CAAA)n#Simple_repeat 209-401, >B3A#SINE/B2
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES            Location/Qualifiers
     source         1. .495
                    /organism="Rattus norvegicus"
                    /strain="Sprague-Dawley"
                    /db_xref="taxon:10116"
                    /clone="UI-R-CA0-bfm-e-11-0-UI"
                    /clone_lib="UI-R-CA0"
                    /lab_host="DH10B (Life Technologies)"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
                    library is a subtracted library derived from the following
                    tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                    , midbrain,  cerebral cortex, corpus striatum, testis, and
                    hippocampus.  For a detailed description of the library
                    from which this clone was  derived, please visit our web
                    site at ratest.eng.uiowa.edu.  The subtraction has been
                    previously described in (Bonaldo,  Lennon and Soares,
                    Genome Research 6:791-806, 1996)
                    TAG_LIB=UI-R-CA0
                    TAG_TISSUE=corpus-striatum
                    TAG_SEQ=CTAGG"
BASE COUNT      162 a    120 c     76 g    137 t
ORIGIN
```

```
  Query Match             9.5%;  Score 37.8;  DB 10;  Length 495;
  Best Local Similarity   52.2%;  Pred. No. 5.6;
  Matches   84;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     36 acatctgaattacaatagatgtgctaagatggtatgtttcggtgtgtattctttggttgc 95
          |||    || |||| | | |    |  | |  | ||||||  ||| | ||  ||  | |
Db    166 ACAGAATAAATACATTGGCTCCTGTTTGTTTCTTTGTTTCTTTGTTTGTTTGTTTTTACC 107

Qy     96 aagcaccaggcatgagttttggtgaatttaatcaagagagtaattcgttacaagaacttg 155
          |      |  || | |||||| ||  |||| |      || | |  ||   | |    ||
Db    106 ATAATTGATTCACGGGTTTTGTTGTTTTTATTTTTTTAAGAATTATGTATTACGGGGCTG 47

Qy    156 gggtttctgtgcaacccaagaaaagtctgaaaataaccaaa 196
          ||| ||  |  ||     | | || ||| |||| ||  |||
Db     46 GGGATTTAGCTCAGTGGTACATAATTCTTAAAAAAAAAAAA 6


RESULT 11
BE146756/c
LOCUS       BE146756                340 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  QV4-HT0222-181099-013-a09 HT0222 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE146756
VERSION     BE146756.1  GI:8609480
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 340)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV4-HT0222-181
            099-013-a09&t3=1999-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 14
            High quality sequence stop: 340.
FEATURES            Location/Qualifiers
     source         1. .340
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="HT0222"
                    /dev_stage="Adult"
                    /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                    Site_2: SmaI; A mini-library was made by cloning products
                    derived from  ORESTES PCR (U.S. Letters Patent application
                    No. 196,716 -  Ludwig Institute for Cancer Research)
                    profiles into  the pUC 18 vector. Reverse transcription of
                    tissue mRNA and cDNA amplification were performed under
                    low stringency conditions."
BASE COUNT       99 a     79 c     96 g     66 t
ORIGIN


  Query Match             9.5%;  Score 37.6;  DB 9;  Length 340;
```

```
  Best Local Similarity   54.3%;  Pred. No. 6;
  Matches   76;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy    191 accaaatttagaaaatgtaaggaatcaggggggtttcaagactctagatagcaagaagtgg 250
          | | |||||    | |   || |   ||||||||||  | | | | |||   ||  | |
Db    235 AGCTAATTTTTGTATTTGTAGTAGAGAGGGGGTTTCACCATTTTGGCTAGGCTGATCTTG 176

Qy    251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcca 310
          ||   |  | ||| || || || ||      |    ||     |   ||||| |  ||
Db    175 AACTCCTGGCCTCAAGCGATCCACCCGCCTTGGCCTTCCAAAGCTCTGGGTTTACAGGCA 116

Qy    311 agtgccattctccccaacct 330
          | |||| | | |||| |||
Db    115 TGAGCCACTGTGCCCAGCCT 96


RESULT 12
BE147023/c
LOCUS       BE147023                341 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  QV4-HT0222-091199-024-b09 HT0222 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE147023
VERSION     BE147023.1  GI:8609747
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV4-HT0222-091
            199-024-b09&t3=1999-11-09&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 341.
FEATURES             Location/Qualifiers
     source          1. .341
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0222"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       99 a     75 c     99 g     68 t
ORIGIN


  Query Match             9.5%;  Score 37.6;  DB 9;  Length 341;
  Best Local Similarity   54.3%;  Pred. No. 6;
  Matches   76;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy    191 accaaatttagaaaatgtaaggaatcaggggggtttcaagactctagatagcaagaagtgg 250
          | | |||||    | |   || |   ||||||||||  | | | | |||   ||  | |
Db    236 AGCTAATTTTTGTATTTGTAGTAGAGAGGGGGTTTCACCATTTTGGCTAGGCTGATCTTG 177

Qy    251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcca 310
          ||   |  | ||| || || || ||      |    ||     |   ||||| |  ||
Db    176 AACTCCTGGCCTCAAGCGATCCACCCGCCTTGGCCTTCCAAAGCTCTGGGTTTACAGGCA 117

Qy    311 agtgccattctccccaacct 330
          | |||| | | |||| |||
Db    116 TGAGCCACTGTGCCCAGCCT 97


RESULT 13
BE146946/c
LOCUS       BE146946                342 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  QV4-HT0222-011199-018-g09 HT0222 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE146946
VERSION     BE146946.1  GI:8609670
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 342)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV4-HT0222-011
            199-018-g09&t3=1999-11-01&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 19
            High quality sequence stop: 342.
FEATURES             Location/Qualifiers
     source          1. .342
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0222"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       98 a     78 c     100 g     65 t      1 others
ORIGIN


  Query Match             9.5%;  Score 37.6;  DB 9;  Length 342;
  Best Local Similarity   54.3%;  Pred. No. 6;
  Matches   76;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy    191 accaaatttagaaaatgtaaggaatcaggggggtttcaagactctagatagcaagaagtgg 250
```

```
          | | |||||    | |   || |   ||||||||||  | | | | |||   ||  | |
Db    237 AGCTAATTTTTGTATTTGTAGTAGAGAGGGGGTTTCACCATTTTGGCTAGGCTGATCTTG 178

Qy    251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcca 310
          ||   |  | ||| || || || ||       |    ||     |    |||||| |  ||
Db    177 AACTCCTGGCCTCAAGCGATCCACCCGCCTTGGCCTTCCAAAGCTCTGGGTTTACAGGCA 118

Qy    311 agtgccattctccccaacct 330
            | |||| | | |||| |||
Db    117 TGAGCCACTGTGCCCAGCCT 98


RESULT 14
AV722572/c
LOCUS       AV722572                 702 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV722572 HTB Homo sapiens cDNA clone HTBANE07 5', mRNA sequence.
ACCESSION   AV722572
VERSION     AV722572.1  GI:10825195
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTB clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HTBANE07"
                     /clone_lib="HTB"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      149 a    184 c    154 g    215 t
ORIGIN


  Query Match             9.5%;  Score 37.6;  DB 9;  Length 702;
  Best Local Similarity   50.6%;  Pred. No. 6.7;
  Matches   91;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy    117 gtgaatttaatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaaga 176
          ||||| ||||   |||   |  ||| |  | || |  |||||  |        ||||  ||
Db    312 GTGAAATTAAAGGAGATCATGCTTCTTACCCAGGAAGTGGGGAATTCAGTTGACCCGTGA 253

Qy    177 aaagtctgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaagactctag 236
          |||  ||||| | |   | ||   ||||   |||  |    |      | |  |  |  ||
Db    252 AAAAACTGAATAAATTAACATGGGGAAAGGCTAATTACCATGCCCAGTGCTGGGGTGGAG 193

Qy    237 atagcaagaagtggaaaatcctgtctccagggaacctcctcaggggagaatcagctccag 296
          | |   ||  |   |   |   |   || ||| |||| || |||    | |||||   ||
Db    192 AAACAGAGTGGCCAAGGTTGTAGCAGCCTGGGCACCTGCTGAGGCCCCATTCAGCGGGAG 133


RESULT 15
BE146803/c
LOCUS       BE146803                 330 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  QV4-HT0222-181099-013-d07 HT0222 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE146803
VERSION     BE146803.1  GI:8609527
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QV4-HT0222-181
            099-013-d07&t3=1999-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 330.
FEATURES             Location/Qualifiers
     source          1. .330
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0222"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       98 a     75 c     95 g     62 t
ORIGIN


  Query Match             9.4%;  Score 37.2;  DB 9;  Length 330;
  Best Local Similarity   51.2%;  Pred. No. 7.7;
  Matches   87;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy    191 accaaatttagaaaatgtaaggaatcaggggtttcaagactctagatagcaagaagtgg 250
          | | |||||    | |   ||      ||||||||||  | | | | |||   ||  | |
Db    225 AGCTAATTTTTGTATTTGTAGTGGAGAGGGGGTTTCACCATTTTGGCTAGGCTGATCTTG 166

Qy    251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcca 310
          ||   |  | ||| || || || ||       |    ||     |    |||||| |  ||
Db    165 AACTCCTGGCCTCAAGCGATCCACCCGCCTTGGCCTTCCAAAGCTCTGGGTTTACAGGCA 106

Qy    311 agtgccattctccccaacctccaggagtgagagtttctttagtatgtctt 360
          | |||| | | |||| |||      | |    || | |  | ||| ||
Db    105 TGAGCCACTGTGCCCAGCCTGTTTTTTTTATTTGTTTTGTTTTTTGTTTT 56


Search completed: August 26, 2002, 23:33:24
Job time: 5884 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2002, 23:11:43 ; Search time 280.27 Seconds
(without alignments)
3923.452 Million cell updates/sec

Title: US-09-785-276A-32
Perfect score: 396
Sequence: 1 ggcgaattggagctccccgc.....ttgtgaccagggcaagaggg 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1757194 seqs, 1388416550 residues

Total number of hits satisfying chosen parameters: 3514388

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq2:*
8: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	396	100.0	396	5	US-09-785-276A-32	Sequence 32, Appl
2	368.4	93.0	427	5	US-09-785-276A-30405	Sequence 30405, A
3	368.4	93.0	427	5	US-09-785-276A-39380	Sequence 39380, A
4	351.4	88.7	477	5	US-09-785-276A-9201	Sequence 9201, Ap
5	48.4	12.2	549	6	US-10-198-846-7591	Sequence 7591, Ap
6	47	11.9	523	5	US-09-785-276A-8342	Sequence 8342, Ap
7	46.6	11.8	480	6	US-10-198-846-1504	Sequence 1504, Ap
8	46.4	11.7	454	5	US-09-785-276A-515	Sequence 515, App
9	46.4	11.7	504	6	US-10-198-846-1516	Sequence 1516, Ap
10	46.2	11.7	858	6	US-10-198-846-3335	Sequence 3335, Ap
11	45.8	11.6	410	5	US-09-785-276A-3202	Sequence 3202, Ap
12	45.8	11.6	446	5	US-09-785-276A-5029	Sequence 5029, Ap
13	45.6	11.5	407	5	US-09-785-276A-3821	Sequence 3821, Ap
14	45.6	11.5	439	5	US-09-785-276A-2133	Sequence 2133, Ap
15	45.2	11.4	244	5	US-09-785-276A-739	Sequence 739, App
16	45	11.4	447	6	US-10-198-846-1370	Sequence 1370, Ap
17	44.8	11.3	458	5	US-09-785-276A-495	Sequence 495, App
18	44.6	11.3	492	6	US-10-198-846-1987	Sequence 1987, Ap
19	44.4	11.2	409	5	US-09-785-276A-1180	Sequence 1180, Ap
20	44.2	11.2	385	5	US-09-785-276A-318	Sequence 318, App
21	44.2	11.2	407	5	US-09-785-276A-1000	Sequence 1000, Ap
22	44.2	11.2	852	6	US-10-198-846-6037	Sequence 6037, Ap
23	44.2	11.2	868	6	US-10-198-846-5959	Sequence 5959, Ap
24	44	11.1	383	6	US-10-198-846-1760	Sequence 1760, Ap
25	44	11.1	476	6	US-10-198-846-1432	Sequence 1432, Ap
26	43.8	11.1	427	5	US-09-785-276A-8752	Sequence 8752, Ap
27	43.8	11.1	642	7	US-10-027-632-280816	Sequence 280816,
28	43.8	11.1	670	5	US-09-785-276A-1152	Sequence 1152, Ap
29	43.6	11.0	439	5	US-09-785-276A-477	Sequence 477, App
30	43.6	11.0	440	5	US-09-785-276A-4349	Sequence 4349, Ap
31	43.6	11.0	452	6	US-10-198-846-7840	Sequence 7840, Ap
32	43.6	11.0	454	6	US-10-198-846-583	Sequence 583, App
33	43.6	11.0	474	6	US-10-198-846-664	Sequence 664, App
34	43.6	11.0	865	6	US-10-198-846-1856	Sequence 1856, Ap
35	43.6	11.0	876	6	US-10-198-846-4121	Sequence 4121, Ap
36	43.4	11.0	325	5	US-09-785-276A-8310	Sequence 8310, Ap
37	43.4	11.0	441	5	US-09-785-276A-4348	Sequence 4348, Ap
38	43.2	10.9	255	5	US-09-785-276A-5191	Sequence 5191, Ap
39	43.2	10.9	417	5	US-09-785-276A-6255	Sequence 6255, Ap
40	43.2	10.9	464	6	US-10-198-846-1125	Sequence 1125, Ap
41	43	10.9	304	5	US-09-785-276A-8426	Sequence 8426, Ap
42	42.8	10.8	253	5	US-09-785-276A-4006	Sequence 4006, Ap
43	42.8	10.8	363	5	US-09-785-276A-4022	Sequence 4022, Ap
44	42.8	10.8	380	5	US-09-785-276A-4091	Sequence 4091, Ap
45	42.8	10.8	437	6	US-10-198-846-1411	Sequence 1411, Ap

ALIGNMENTS

RESULT 1
US-09-785-276A-32
; Sequence 32, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel. Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007B
; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-785-276A-32

Query Match 100.0%; Score 396; DB 5; Length 396;
Best Local Similarity 100.0%; Pred. No. 1.8e-100;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60

Qy      61 aagatggtatgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtga 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 aagatggtatgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtga 120

Qy     121 atttaatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    121 atttaatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaag 180

Qy    181 tctgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaagactctagatag 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 tctgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaagactctagatag 240

Qy    241 caagaagtggaaaatcctgtctccagggaacctcctcaggggagaatcagctccagctgg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 caagaagtggaaaatcctgtctccagggaacctcctcaggggagaatcagctccagctgg 300

Qy    301 tttatatccaagtgccattctccccaacctccaggagtgagagtttctttagtatgtctt 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 tttatatccaagtgccattctccccaacctccaggagtgagagtttctttagtatgtctt 360

Qy    361 ggatcctatgcccactttgtgaccagggcaagaggg 396
          |||||||||||||||||||||||||||||||||||
Db    361 ggatcctatgcccactttgtgaccagggcaagaggg 396


RESULT   2
US-09-785-276A-30405
; Sequence 30405, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30405
;   LENGTH: 427
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-785-276A-30405


  Query Match             93.0%;  Score 368.4;  DB 5;  Length 427;
  Best Local Similarity   99.2%;  Pred. No. 8.8e-93;
  Matches  391;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy      4 gaattggagctcc-ccgcggtggcggccgaggtacatctgaattacaatagatgtgctaa 62
          |||||||||||||  |||||||||||||||||||||||||||||||| ||||||||||||
Db      1 gaattggagctccaccgcggtggcggccgaggtacatctgaattac-atagatgtgctaa 59

Qy     63 gatggtatgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaat 122
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 gatggtatgtgtcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaat 119

Qy    123 ttaatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtc 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 ttaatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtc 179

Qy    183 tgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaagactctagatagca 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 tgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaagactctagatagca 239

Qy    243 agaagtggaaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtt 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 agaagtggaaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtt 299

Qy    303 tatatccaagtgccattctccccaacctccaggagtgagagtttctttagtatgtcttgg 362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 tatatccaagtgccattctccccaacctccaggagtgagagtttctttagtatgtcttgg 359

Qy    363 atcctatgcccactttgtgaccagggcaagaggg 396
          |||||||||||||||||||||||||||||||||||
Db    360 atcctatgcccactttgtgaccagggcaagaggg 393


RESULT   3
US-09-785-276A-39380
; Sequence 39380, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39380
;   LENGTH: 427
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-785-276A-39380


  Query Match             93.0%;  Score 368.4;  DB 5;  Length 427;
  Best Local Similarity   99.2%;  Pred. No. 8.8e-93;
  Matches  391;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy      4 gaattggagctcc-ccgcggtggcggccgaggtacatctgaattacaatagatgtgctaa 62
          |||||||||||||  |||||||||||||||||||||||||||||||| ||||||||||||
Db      1 gaattggagctccaccgcggtggcggccgaggtacatctgaattac-atagatgtgctaa 59

Qy     63 gatggtatgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaat 122
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 gatggtatgtgtcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaat 119

Qy    123 ttaatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtc 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 ttaatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtc 179

Qy    183 tgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaagactctagatagca 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 tgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaagactctagatagca 239

Qy    243 agaagtggaaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtt 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 agaagtggaaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtt 299

Qy    303 tatatccaagtgccattctccccaacctccaggagtgagagtttctttagtatgtcttgg 362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS   Claim 1; Page 494; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 320 BP; 77 A; 79 C; 71 G; 73 T; 20 other;

  Query Match           11.8%;  Score 46.6;  DB 22;  Length 320;
  Best Local Similarity  74.4%;  Pred. No. 0.00011;
  Matches   58;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
           ||||||||||||||||||||||||||||||||||||||||||   ||     |||| ||| || |
Db       9 ggcgaattggagctccccgcggtggcggccgaggtaccattggtggccaattgatttgat 68

Qy      61 aagatggtatgtttcggt 78
               | || | |  ||| |
Db      69 ggtaagggagggatcgnt 86


RESULT   2
AAL10167
ID   AAL10167 standard; cDNA; 377 BP.
XX
AC   AAL10167;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 2624.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS   Claim 1; Page 496; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 377 BP; 87 A; 62 C; 87 G; 116 T; 25 other;

  Query Match           11.7%;  Score 46.4;  DB 22;  Length 377;
  Best Local Similarity  64.8%;  Pred. No. 0.00013;
  Matches   68;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
           |||||||||||||||||||||||||||||||||||||||| | |   ||    |  | | |
Db      10 ggcgaattggagctccccgcggtggcggccgaggtacttttttttttttttttttttttt 69

Qy      61 aagatggtatgtttcggtgtgtattctttggttgcaagcaccagg 105
             |   | | |||   | | | || |||  || ||  ||  |||
Db      70 tttaatttttttttttttttttttttttttttttttncactcagaagg 114


RESULT   3
AAL11997
ID   AAL11997 standard; cDNA; 517 BP.
XX
AC   AAL11997;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 4454.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS   Claim 1; Page 805; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 517 BP; 201 A; 94 C; 102 G; 117 T; 3 other;
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2002, 22:32:26 ; Search time 210.62 Seconds
(without alignments)
3228.080 Million cell updates/sec

Title: US-09-785-276A-32
Perfect score: 396
Sequence: 1 ggcgaattggagctccccgc.....ttgtgaccagggcaagaggg 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	46.6	11.8	320	22	AAL10155	Human breast cance
	2	46.4	11.7	377	22	AAL10167	Human breast cance
	3	46.2	11.7	517	22	AAL11997	Human breast cance
	4	45	11.4	472	22	AAL08312	Human breast cance
	5	44	11.1	247	22	AAL10083	Human breast cance
	6	43.6	11.0	447	22	AAL09232	Human breast cance
	7	43.6	11.0	555	22	AAL12840	Human breast cance
	8	43.2	10.9	423	22	AAL07597	Human breast cance
	9	43.2	10.9	449	22	AAL09775	Human breast cance
	10	42.8	10.8	128	22	AAL07701	Human breast cance
	11	42.8	10.8	247	22	AAL11111	Human breast cance
	12	42.8	10.8	257	22	AAL11039	Human breast cance
	13	42.8	10.8	312	22	AAL08106	Human breast cance
	14	42.8	10.8	403	22	AAL10062	Human breast cance
c	15	42.8	10.8	1167	22	AAL26699	Human breast cance
	16	42.6	10.8	431	22	AAL10341	Human breast cance
	17	42.4	10.7	256	22	AAL10126	Human breast cance
	18	42.4	10.7	368	22	AAL08243	Human breast cance
	19	42	10.6	375	22	AAL10505	Human breast cance
	20	42	10.6	427	22	AAL11652	Human breast cance
	21	42	10.6	433	22	AAL10669	Human breast cance
	22	42	10.6	466	22	AAL09253	Human breast cance
	23	42	10.6	469	22	AAL09264	Human breast cance
	24	42	10.6	610	22	AAL10486	Human breast cance
	25	42	10.6	705	22	AAL10377	Human breast cance
	26	41.8	10.6	186	22	AAL10266	Human breast cance
	27	41.8	10.6	353	22	AAL11349	Human breast cance
	28	41.8	10.6	409	22	AAL08417	Human breast cance
	29	41.8	10.6	465	22	AAL11238	Human breast cance
	30	41.8	10.6	524	22	AAL07911	Human breast cance
	31	41.6	10.5	363	22	AAL10263	Human breast cance
	32	41.6	10.5	444	22	AAL07993	Human breast cance
	33	41.6	10.5	470	22	AAL09075	Human breast cance
	34	41.6	10.5	534	22	AAL07776	Human breast cance
	35	41.6	10.5	785	22	AAL12395	Human breast cance
	36	41.4	10.5	141	22	AAL07967	Human breast cance
	37	41.4	10.5	234	22	AAL10172	Human breast cance
	38	41.4	10.5	252	22	AAL09107	Human breast cance
	39	41.4	10.5	390	22	AAL09561	Human breast cance
	40	41.4	10.5	391	22	AAL10261	Human breast cance
	41	41.4	10.5	403	22	AAL09538	Human breast cance
	42	41.4	10.5	469	22	AAL10106	Human breast cance
	43	41.4	10.5	498	22	AAL09090	Human breast cance
	44	41.4	10.5	528	22	AAL11295	Human breast cance
	45	41.4	10.5	741	22	AAL11654	Human breast cance

ALIGNMENTS

RESULT 1
AAL10155
ID AAL10155 standard; cDNA; 320 BP.
XX
AC AAL10155;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 2612.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX

```
Db    300 tatatccaagtgccattctccccaacctccaggagtgagagtttctttagtatgtcttgg 359

Qy    363 atcctatgcccactttgtgaccagggcaagaggg 396
          |||||||||||||||||||||||||||||||||
Db    360 atcctatgcccactttgtgaccagggcaagaggg 393


RESULT   4
US-09-785-276A-9201
; Sequence 9201, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9201
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-785-276A-9201


  Query Match             88.7%;  Score 351.4;  DB 5;  Length 477;
  Best Local Similarity   99.5%;  Pred. No. 4.9e-88;
  Matches  363;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy     32 aggtacatctgaattacaatagatgtgctaagatggtatgtttcggtgtgtattctttgg 91
          |||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||
Db      1 aggtacatctgaattac-atagatgtgctaagatggtatgtgtcggtgtgtattctttgg 59

Qy     92 ttgcaagcaccaggcatgagttttggtgaatttaatcaagagagtaattcgttacaagaa 151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 ttgcaagcaccaggcatgagttttggtgaatttaatcaagagagtaattcgttacaagaa 119

Qy    152 cttggggtttctgtgcaacccaagaaaagtctgaaaataaccaaatttagaaaatgtaag 211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 cttggggtttctgtgcaacccaagaaaagtctgaaaataaccaaatttagaaaatgtaag 179

Qy    212 gaatcaggggtttcaagactctagatagcaagaagtggaaaatcctgtctccagggaac 271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 gaatcaggggtttcaagactctagatagcaagaagtggaaaatcctgtctccagggaac 239

Qy    272 ctcctcaggggagaatcagctccagctggtttatatccaagtgccattctccccaacctc 331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 ctcctcaggggagaatcagctccagctggtttatatccaagtgccattctccccaacctc 299

Qy    332 caggagtgagagtttctttagtatgtcttggatcctatgcccactttgtgaccagggcaa 391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 caggagtgagagtttctttagtatgtcttggatcctatgcccactttgtgaccagggcaa 359

Qy    392 gaggg 396
          |||||
Db    360 gaggg 364


RESULT   5
US-10-198-846-7591
; Sequence 7591, Application US/10198846
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7591
;   LENGTH: 549
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2, 3, 4, 8, 9, 114, 494, 502, 506, 511, 512, 525, 527, 530
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-7591


  Query Match             12.2%;  Score 48.4;  DB 6;  Length 549;
  Best Local Similarity   78.4%;  Pred. No. 0.0013;
  Matches   58;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          |||||||||||||||||||||||||||||||||||||| |   |    ||| || | ||
Db     18 ggcgaattggagctccccgcggtggcggccgaggtacatatatgtgtgtatatatatact 77

Qy     61 aagatggtatgttt 74
          | |||  || |||
Db     78 taaatgtaatattt 91


RESULT   6
US-09-785-276A-8342
; Sequence 8342, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8342
;   LENGTH: 523
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 4, 328, 442
;   OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-8342


  Query Match             11.9%;  Score 47;  DB 5;  Length 523;
  Best Local Similarity   74.7%;  Pred. No. 0.0031;
  Matches   59;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||||||||||||||||||||||||||||||||||||||||||  ||     |||| ||| || |
Db     11 ggcgaattggagctccccgcggtggcggccgaggtaccattggtggccaattgatttgat 70

Qy     61 aagatggtatgtttcggtg 79
              | || | |  ||| ||
Db     71 ggtaagggagggatcgttg 89


RESULT   7
US-10-198-846-1504
; Sequence 1504, Application US/10198846
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1504
;   LENGTH: 480
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 2, 99, 101, 109, 118, 121, 124, 180, 181, 189, 203, 204,
;   LOCATION: 210, 211, 212, 214, 220, 269, 272, 290, 307, 346, 368, 395,
;   LOCATION: 411, 415, 445, 448, 449, 458, 461, 463, 469, 476
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-1504


  Query Match             11.8%;  Score 46.6;  DB 6;  Length 480;
  Best Local Similarity   74.4%;  Pred. No. 0.004;
  Matches   58;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||||||||||||||||||||||||||||||||||||  ||     |||| ||| || |
Db     23 ggcgaattggagctccccgcggtggcggccgaggtaccattggtggccaattgatttgat 82

Qy     61 aagatggtatgtttcggt 78
              | || | |  ||| |
Db     83 ggtaagggagggatcgnt 100


RESULT   8
US-09-785-276A-515
; Sequence 515, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 515
;   LENGTH: 454
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 6, 353, 404, 421, 430
;   OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-515


  Query Match             11.7%;  Score 46.4;  DB 5;  Length 454;
  Best Local Similarity   70.4%;  Pred. No. 0.0045;
  Matches   76;  Conservative    0;  Mismatches   31;  Indels    1;  Gaps    1;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||||||||||||||||||||||||||||||||||||||||| ||| ||||| |||  | | |||
Db     12 ggcgaattggagctccccgcggtggcggccgaggtacttcttaattagaattta-gggct 70

Qy     61 aagatggtatgtttcggtgtgtattctttggttgcaagcaccaggcat 108
          || | |  || ||       |  || |||   ||  |   |  |||| |
Db     71 aacaagagatattataacatcaatacttgcattatatagaggaggcct 118


RESULT   9
US-10-198-846-1516
; Sequence 1516, Application US/10198846
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1516
;   LENGTH: 504
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2, 3, 117, 131, 172, 194, 196, 200, 210, 230, 231, 255, 259,
;   LOCATION: 264, 265, 269, 274, 282, 289, 306, 313, 323, 332, 360, 365,
;   LOCATION: 376, 381, 404, 411, 416, 419, 437, 444, 451, 452, 459, 468,
;   LOCATION: 474, 481, 499
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-1516
```

```
  Query Match             11.7%;  Score 46.4;  DB 6;  Length 504;
  Best Local Similarity   64.8%;  Pred. No. 0.0046;
  Matches   68;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
           ||||||||||||||||||||||||||||||||||| | |   ||    |   | | |
Db      24 ggcgaattggagctccccgcggtggcggccgaggtacttttttttttttttttttttttt 83

Qy      61 aagatggtatgtttcggtgtgtattctttggttgcaagcaccagg 105
             |   | | |||   | | | || |||  || ||  ||  |||
Db      84 tttaatttttttttttttttttttttttttttttncactcagaagg 128


RESULT  10
US-10-198-846-3335
; Sequence 3335, Application US/10198846
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3335
;   LENGTH: 858
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 442, 486, 524, 558, 650, 700, 770, 772, 774, 779, 795, 798,
;   LOCATION: 811, 841, 842, 845, 846, 847, 848, 849, 850, 851, 852, 853,
;   LOCATION: 854, 855
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-3335


  Query Match             11.7%;  Score 46.2;  DB 6;  Length 858;
  Best Local Similarity   80.6%;  Pred. No. 0.006;
  Matches   54;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
           ||||||||||||||||||||||||||||||||||| ||  ||||  || | | | |
Db      24 ggcgaattggagctccccgcggtggcggccgaggtacctcacaatttaaaaaaacgggga 83

Qy      61 aagatgg 67
           || || |
Db      84 aaaatag 90


RESULT  11
US-09-785-276A-3202
; Sequence 3202, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3202
;   LENGTH: 410
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 313
;   OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-3202


  Query Match             11.6%;  Score 45.8;  DB 5;  Length 410;
  Best Local Similarity   81.5%;  Pred. No. 0.0064;
  Matches   53;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||    | |  ||  |||
Db      12 ggcgaattggagctccccgcggtggcggccgaggtacatctggcgcagaccagcagtggc 71

Qy      61 aagat 65
           | |||
Db      72 acgat 76


RESULT  12
US-09-785-276A-5029
; Sequence 5029, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5029
;   LENGTH: 446
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-785-276A-5029


  Query Match             11.6%;  Score 45.8;  DB 5;  Length 446;
  Best Local Similarity   81.5%;  Pred. No. 0.0065;
```

```
  Matches   53;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||||||||||||||||||||||||||||||||||| | |  |||  |||  |    ||
Db      9 ggcgaattggagctccccgcggtggcggccgaggtacttttatattttaattcaaaacct 68

Qy     61 aagat 65
          |||||
Db     69 aagat 73


RESULT  13
US-09-785-276A-3821
; Sequence 3821, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3821
;   LENGTH: 407
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 201, 311, 337, 376, 389
;   OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-3821


  Query Match            11.5%;  Score 45.6;  DB 5;  Length 407;
  Best Local Similarity   75.0%;  Pred. No. 0.0072;
  Matches   57;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||||||||||||||||||||||||||||||||||  || || || | |      |
Db      9 ggcgaattggagctccccgcggtggcggccgaggtacaattgcatcaccaaaagctgttt 68

Qy     61 aagatggtatgtttcg 76
          ||  |||  ||| | |
Db     69 aatttggattgtgtag 84


RESULT  14
US-09-785-276A-2133
; Sequence 2133, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2133
;   LENGTH: 439
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 14, 66, 69, 71, 76, 292, 317, 328, 338, 343, 428
;   OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-2133


  Query Match            11.5%;  Score 45.6;  DB 5;  Length 439;
  Best Local Similarity   74.0%;  Pred. No. 0.0074;
  Matches   54;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          |||||||| |||||||||||||||||||||||||||||||| |||     || || |||  |
Db      5 ggcgaattgnagctccccgcggtggcggccgaggtacatgtgatggtgaaaagttgtttt 64

Qy     61 aagatggtatgtt 73
          |    | | |  |
Db     65 anagngnttttnt 77


RESULT  15
US-09-785-276A-739
; Sequence 739, Application US/09785276A
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007B
;  CURRENT APPLICATION NUMBER: US/09/785,276A
;  CURRENT FILING DATE:  2001-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 739
;   LENGTH: 244
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
```

; LOCATION: 213, 223, 225, 231
; OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-739

Query Match 11.4%; Score 45.2; DB 5; Length 244;
Best Local Similarity 75.7%; Pred. No. 0.0082;
Matches 56; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
           ||||||||||||||||||||||||||||||||||||||    |  ||  ||  | || |
Db       1 ggcgaattggagctccccgcggtggcggccgaggtactcaggctttcaaagggttgagga 60

Qy      61 aagatggtatgttt 74
           ||| |  | |||||
Db      61 aaggtcttttgttt 74
```

Search completed: August 27, 2002, 00:33:20
Job time: 4897 sec

```
  Query Match           11.7%;  Score 46.2;  DB 22;  Length 517;
  Best Local Similarity  80.6%;  Pred. No. 0.00018;
  Matches   54;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||||||||||||||||||||||||||||||||||| ||  ||||  || | | | |
Db      4 ggcgaattggagctccccgcggtggcggccgaggtacctcacaatttaaaaaaacgggga 63

Qy     61 aagatgg 67
          || || |
Db     64 aaaatag 70


RESULT   4
AAL08312
ID   AAL08312 standard; cDNA; 472 BP.
XX
AC   AAL08312;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 769.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 210; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 472 BP; 134 A; 99 C; 91 G; 146 T; 2 other;


  Query Match           11.4%;  Score 45;  DB 22;  Length 472;
  Best Local Similarity  82.3%;  Pred. No. 0.00041;
  Matches   51;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||| ||||||||||||||||||||||||||||||||| || |||  ||| |||    ||
Db      5 ggcnaattggagctccccgcggtggcggccgaggtacagcttaatggcaaaagagaatct 64

Qy     61 aa 62
          |
Db     65 ta 66


RESULT   5
AAL10083
ID   AAL10083 standard; cDNA; 247 BP.
XX
AC   AAL10083;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 2540.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 483; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 247 BP; 61 A; 58 C; 58 G; 69 T; 1 other;


  Query Match           11.1%;  Score 44;  DB 22;  Length 247;
  Best Local Similarity  53.5%;  Pred. No. 0.00064;
  Matches   92;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||||||||||||||||||||||||||||||||||||  | ||||  |       |
Db      9 ggcgaattggagctccccgcggtggcggccgaggtacaaattaatttaacacatcttttg 68

Qy     61 aagatggtatgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtga 120
          |  ||   ||  || |   || |      ||    |  ||  ||     | |||  |
Db     69 ataatctcatccttggtgttggaaaagacgggaaaatccaaaagtgtctattttgtgccc 128

Qy    121 atttaatcaagagagtaattcgttacaagaacttggggtttctgtgcaaccc 172
          |  |  |||||  | || |      |    |  |  ||  |||   || |||
Db    129 aaatgctcaagttaatactcgagggccttgatctctggcctcttgccatccc 180
```

```
RESULT   6
AAL09232
ID   AAL09232 standard; cDNA; 447 BP.
XX
AC   AAL09232;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 1689.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS   Claim 1; Page 338-339; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 447 BP; 117 A; 84 C; 93 G; 153 T; 0 other;

  Query Match             11.0%;  Score 43.6;  DB 22;  Length 447;
  Best Local Similarity   84.5%;  Pred. No. 0.0011;
  Matches   49;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtg 58
           |||||||||||||||||||||||||||||||||||| |     ||   ||||||||
Db      12 ggcgaattggagctccccgcggtggcggccgaggtacagcatggataatttagatgtg 69


RESULT   7
AAL12840
ID   AAL12840 standard; cDNA; 555 BP.
XX
AC   AAL12840;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 5297.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS   Claim 1; Page 951; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 555 BP; 130 A; 148 C; 136 G; 141 T; 0 other;

  Query Match             11.0%;  Score 43.6;  DB 22;  Length 555;
  Best Local Similarity   92.0%;  Pred. No. 0.0012;
  Matches   46;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaa 50
           |||||||||||||||||||||||||||||||||||||||||||| |||   ||
Db       1 ggcgaattggagctccccgcggtggcggccgaggtacatcttaatagaaa 50


RESULT   8
AAL07597
ID   AAL07597 standard; cDNA; 423 BP.
XX
AC   AAL07597;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 54.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
```

PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 114; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 423 BP; 117 A; 68 C; 62 G; 176 T; 0 other;

Query Match 10.9%; Score 43.2; DB 22; Length 423;
Best Local Similarity 68.2%; Pred. No. 0.0014;
Matches 60; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
||||||||||||||||||||||||||||||||||| || | || ||||
Db 5 ggcgaattggagctccccgcggtggcggccgaggtacctcacgcgcataaatttgctgct 64

Qy 61 aagatggtatgtttcggtgtgtattctt 88
|| | | | | ||| || ||
Db 65 cctatttttttttctgtttatgtgttttt 92

RESULT 9
AAL09775
ID AAL09775 standard; cDNA; 449 BP.
XX
AC AAL09775;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 2232.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 429; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 449 BP; 87 A; 129 C; 94 G; 131 T; 8 other;

Query Match 10.9%; Score 43.2; DB 22; Length 449;
Best Local Similarity 71.2%; Pred. No. 0.0015;
Matches 57; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
|||||||||||||||||||||||||||||||||| | | || | | | |
Db 8 ggcgaattggagctccccgcggtggcggccgaggtacttttttttttttttttttttttt 67

Qy 61 aagatggtatgtttcggtgt 80
||| | || || | |||
Db 68 gagacaggatcttactttgt 87

RESULT 10
AAL07701
ID AAL07701 standard; cDNA; 128 BP.
XX
AC AAL07701;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 158.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 128; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and

```
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 128 BP; 41 A; 23 C; 28 G; 32 T; 4 other;


  Query Match             10.8%;  Score 42.8;  DB 22;  Length 128;
  Best Local Similarity   89.8%;  Pred. No. 0.0012;
  Matches   44;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattaca 49
          |||||||||||||||||||||||||||||||||||||||||   ||| | |
Db      9 ggcgaattggagctccccgcggtggcggccgaggtacatnnaaatnata 57


RESULT  11
AAL11111
ID   AAL11111 standard; cDNA; 247 BP.
XX
AC   AAL11111;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 3568.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS   Claim 1; Page 657; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 247 BP; 68 A; 69 C; 62 G; 48 T; 0 other;


  Query Match             10.8%;  Score 42.8;  DB 22;  Length 247;
  Best Local Similarity   87.0%;  Pred. No. 0.0015;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaataga 54
          |||||||||||||||||||||||||||||||||||||||||| |||||  | |   ||
Db      4 ggcgaattggagctccccgcggtggcggccgaggtacaactgaagcaaagagga 57


RESULT  12
AAL11039
ID   AAL11039 standard; cDNA; 257 BP.
XX
AC   AAL11039;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 3496.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS   Claim 1; Page 643-644; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 257 BP; 71 A; 74 C; 62 G; 50 T; 0 other;


  Query Match             10.8%;  Score 42.8;  DB 22;  Length 257;
  Best Local Similarity   87.0%;  Pred. No. 0.0016;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaataga 54
          |||||||||||||||||||||||||||||||||||||||||| |||||  | |   ||
Db     13 ggcgaattggagctccccgcggtggcggccgaggtacaactgaagcaaagagga 66


RESULT  13
AAL08106
ID   AAL08106 standard; cDNA; 312 BP.
XX
AC   AAL08106;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 563.
XX
```

```
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 185; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 312 BP; 108 A; 63 C; 71 G; 67 T; 3 other;


  Query Match             10.8%;  Score 42.8;  DB 22;  Length 312;
  Best Local Similarity   74.6%;  Pred. No. 0.0017;
  Matches   53;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||||||||||||||||||||||||||||||||||||| ||| |    || |
Db      7 ggcgaattggagctccccgcggtggcggccgaggtacatcggaaattatctacaatgaag 66

Qy     61 aagatggtatg 71
          || ||    ||
Db     67 aaaatttnttg 77


RESULT  14
AAL10062
ID   AAL10062 standard; cDNA; 403 BP.
XX
AC   AAL10062;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 2519.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 479-480; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 403 BP; 106 A; 80 C; 118 G; 96 T; 3 other;


  Query Match             10.8%;  Score 42.8;  DB 22;  Length 403;
  Best Local Similarity   95.7%;  Pred. No. 0.0019;
  Matches   44;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaatt 46
          |||||||||||||||||||||||||||||||||||| | ||||||
Db      6 ggcgaattggagctccccgcggtggcggccgaggtacctgtgaatt 51


RESULT  15
AAL26699/c
ID   AAL26699 standard; cDNA; 1167 BP.
XX
AC   AAL26699;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 19156.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
```

DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 3645; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 1167 BP; 298 A; 246 C; 295 G; 327 T; 1 other;

Query Match 10.8%; Score 42.8; DB 22; Length 1167;
Best Local Similarity 75.7%; Pred. No. 0.003;
Matches 53; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy      1 ggcgaattggagctccccgcggtggcggccgaggtacatctgaattacaatagatgtgct 60
          ||||||||||||||||||||||||||||||||||||||      || | || |  |||
Db   1164 GGCGAATTGGAGCTCCCCGCGGTGGCGGCCGAGGTACATGCAGGGTATATTACAATTGCA 1105

Qy     61 aagatggtat 70
          |  ||  | |
Db   1104 ATCATTATTT 1095
```

Search completed: August 27, 2002, 00:27:10
Job time: 6884 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2002, 21:26:00 ; Search time 1896.6 Seconds
(without alignments)
4369.350 Million cell updates/sec

Title: US-09-785-276A-32
Perfect score: 396
Sequence: 1 ggcgaattggagctccccgc.....ttgtgaccagggcaagaggg 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		362.2	91.5	143800	2	AC011875	AC011875 Homo sapi
2	c	362.2	91.5	186962	9	AP000848	AP000848 Homo sapi
3		360.6	91.1	115762	2	AP000663	AP000663 Homo sapi
4		348.6	88.0	172830	2	AP001320	AP001320 Homo sapi
5	c	45.4	11.5	45029	2	AC017280	AC017280 Drosophil
6		45.4	11.5	68626	2	AC006168	AC006168 Drosophil
7	c	45.4	11.5	102621	3	DMBR43E12	AL138971 Drosophil
8	c	45.4	11.5	181443	3	AC104148	AC104148 Drosophil
9	c	45.4	11.5	300205	3	AE003425	AE003425 Drosophil
10	c	44.6	11.3	161123	9	AC011029	AC011029 Homo sapi
11		44.2	11.2	150152	2	AL355502	AL355502 Homo sapi
12		43	10.9	113559	9	AC025578	AC025578 Homo sapi
13	c	42.8	10.8	159229	9	AL359180	AL359180 Human DNA
14		42.6	10.8	113254	9	AL137009	AL137009 Human DNA
15		42.4	10.7	259894	9	HUAC002302	AC002302 Homo sapi
16		41.8	10.6	114412	9	AL365260	AL365260 Human DNA
17	c	41.8	10.6	153046	9	AL590009	AL590009 Human DNA
18	c	41.8	10.6	157698	2	AC025692	AC025692 Homo sapi
19		41.8	10.6	179463	2	AL596453	AL596453 Homo sapi
20	c	41.8	10.6	213294	2	AC099778	AC099778 Homo sapi
21	c	41.4	10.5	154918	9	HSDJ67A8	AL121957 Human DNA
22		41.4	10.5	155362	2	AC022244	AC022244 Homo sapi
23		41.2	10.4	410	6	AX284391	AX284391 Sequence
24		41.2	10.4	166095	9	AC006052	AC006052 Homo sapi
25	c	41	10.4	171640	30	AC055875	Ac055875 Homo sapi
26		41	10.4	186773	2	AC106900	AC106900 Homo sapi
27		40.8	10.3	306	6	AX284580	AX284580 Sequence
28	c	40.2	10.2	99011	2	AC004937	AC004937 Homo sapi
29	c	40.2	10.2	110000	9	AC005079_0	AC005079 Homo sapi
30		40.2	10.2	161898	2	AC021683	AC021683 Homo sapi
31	c	40.2	10.2	164950	2	AC105191	AC105191 Homo sapi
32		40	10.1	182	6	AX284268	AX284268 Sequence
33		40	10.1	71474	9	AC098819	AC098819 Homo sapi
34		40	10.1	161872	2	AC015949	AC015949 Homo sapi
35		39.8	10.1	42418	9	AL445990	AL445990 Human DNA
36	c	39.8	10.1	186102	2	AC015905	AC015905 Homo sapi
37		39.6	10.0	89837	2	AC091822	AC091822 Homo sapi
38	c	39.6	10.0	122050	2	AC091512	AC091512 Leishmani
39		39.6	10.0	133715	3	AC087838	AC087838 Leishmani
40	c	39.6	10.0	204767	2	AC011367	AC011367 Homo sapi
41	c	39.6	10.0	251664	2	AC008533	AC008533 Homo sapi
42		39.4	9.9	633	6	AX284552	AX284552 Sequence
43	c	39.4	9.9	127593	9	AC006236	AC006236 Homo sapi
44		39.4	9.9	161920	9	AC025169	AC025169 Homo sapi
45		39.4	9.9	166895	2	AC023415	AC023415 Homo sapi

ALIGNMENTS

RESULT 1
AC011875
LOCUS AC011875 143800 bp DNA linear HTG 12-MAR-2000
DEFINITION Homo sapiens clone RP11-16K5, WORKING DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION AC011875
VERSION AC011875.3 GI:7107950
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143800)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-16K5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 143800)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

```
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6453961.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3566
                Center clone name: 16_K_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 93387 bases at least Q40
                Consensus quality: 115701 bases at least Q30
                Consensus quality: 130381 bases at least Q20
                Insert size: 141000; agarose-fp
                Insert size: 141300; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1029: contig of 1029 bp in length
            *     1030 1129: gap of      100 bp
            *     1130     2964: contig of 1835 bp in length
            *     2965 3064: gap of      100 bp
            *     3065     4253: contig of 1189 bp in length
            *     4254 4353: gap of      100 bp
            *     4354     5678: contig of 1325 bp in length
            *     5679 5778: gap of      100 bp
            *     5779     6855: contig of 1077 bp in length
            *     6856 6955: gap of      100 bp
            *     6956     8847: contig of 1892 bp in length
            *     8848 8947: gap of      100 bp
            *     8948    10945: contig of 1998 bp in length
            *    10946 11045: gap of      100 bp
            *    11046    13001: contig of 1956 bp in length
            *    13002 13101: gap of      100 bp
            *    13102    15430: contig of 2329 bp in length
            *    15431 15530: gap of      100 bp
            *    15531    17919: contig of 2389 bp in length
            *    17920 18019: gap of      100 bp
            *    18020    20135: contig of 2116 bp in length
            *    20136 20235: gap of      100 bp
            *    20236    23750: contig of 3515 bp in length
            *    23751 23850: gap of      100 bp
            *    23851    26794: contig of 2944 bp in length
            *    26795 26894: gap of      100 bp
            *    26895    29482: contig of 2588 bp in length
            *    29483 29582: gap of      100 bp
            *    29583    33174: contig of 3592 bp in length
            *    33175 33274: gap of      100 bp
            *    33275    35185: contig of 1911 bp in length
            *    35186 35285: gap of      100 bp
            *    35286    39745: contig of 4460 bp in length
            *    39746 39845: gap of      100 bp
            *    39846    44222: contig of 4377 bp in length
            *    44223 44322: gap of      100 bp
            *    44323    48990: contig of 4668 bp in length
            *    48991 49090: gap of      100 bp
            *    49091    57790: contig of 8700 bp in length
            *    57791 57890: gap of      100 bp
            *    57891    66822: contig of 8932 bp in length
            *    66823 66922: gap of      100 bp
            *    66923    76709: contig of 9787 bp in length
            *    76710 76809: gap of      100 bp
            *    76810    92865: contig of 16056 bp in length
            *    92866 92965: gap of      100 bp
            *    92966   106278: contig of 13313 bp in length
            *   106279 106378: gap of      100 bp
            *   106379   123041: contig of 16663 bp in length
            *   123042 123141: gap of      100 bp
            *   123142   143800: contig of 20659 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143800
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-16K5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1029
                     /note="assembly_fragment"
     misc_feature    1130. .2964
                     /note="assembly_fragment"
     misc_feature    3065. .4253
                     /note="assembly_fragment"
     misc_feature    4354. .5678
                     /note="assembly_fragment"
     misc_feature    5779. .6855
                     /note="assembly_fragment"
     misc_feature    6956. .8847
                     /note="assembly_fragment"
     misc_feature    8948. .10945
                     /note="assembly_fragment"
     misc_feature    11046. .13001
                     /note="assembly_fragment"
     misc_feature    13102. .15430
                     /note="assembly_fragment"
     misc_feature    15531. .17919
                     /note="assembly_fragment"
     misc_feature    18020. .20135
                     /note="assembly_fragment"
     misc_feature    20236. .23750
                     /note="assembly_fragment"
     misc_feature    23851. .26794
                     /note="assembly_fragment"
     misc_feature    26895. .29482
                     /note="assembly_fragment"
     misc_feature    29583. .33174
                     /note="assembly_fragment"
     misc_feature    33275. .35185
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    35286. .39745
                     /note="assembly_fragment"
     misc_feature    39846. .44222
                     /note="assembly_fragment"
     misc_feature    44323. .48990
                     /note="assembly_fragment"
     misc_feature    49091. .57790
                     /note="assembly_fragment"
     misc_feature    57891. .66822
                     /note="assembly_fragment"
     misc_feature    66923. .76709
                     /note="assembly_fragment"
```

```
     misc_feature    76810. .92865
                     /note="assembly_fragment"
     misc_feature    92966. .106278
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    106379. .123041
                     /note="assembly_fragment"
     misc_feature    123142. .143800
                     /note="assembly_fragment"
BASE COUNT    43074 a  27702 c  27628 g  42892 t   2504 others
ORIGIN


  Query Match             91.5%;  Score 362.2;  DB 2;  Length 143800;
  Best Local Similarity   99.2%;  Pred. No. 1e-98;
  Matches  364;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      30 cgaggtacatctgaattacaatagatgtgctaagatggtatgtttcggtgtgtattcttt 89
           ||  |||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   80915 CGTAGTACATCTGAATTACAATAGATGTGCTAAGATGGTATGTGTCGGTGTGTATTCTTT 80974

Qy      90 ggttgcaagcaccaggcatgagttttggtgaatttaatcaagagagtaattcgttacaag 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   80975 GGTTGCAAGCACCAGGCATGAGTTTTGGTGAATTTAATCAAGAGAGTAATTCGTTACAAG 81034

Qy     150 aacttggggtttctgtgcaacccaagaaaagtctgaaaataaccaaatttagaaaatgta 209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81035 AACTTGGGGTTTCTGTGCAACCCAAGAAAAGTCTGAAAATAACCAAATTTAGAAAATGTA 81094

Qy     210 aggaatcaggggggtttcaagactctagatagcaagaagtggaaaatcctgtctccaggga 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81095 AGGAATCAGGGGGTTTCAAGACTCTAGATAGCAAGAAGTGGAAAATCCTGTCTCCAGGGA 81154

Qy     270 acctcctcaggggagaatcagctccagctggtttatatccaagtgccattctccccaacc 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81155 ACCTCCTCAGGGGAGAATCAGCTCCAGCTGGTTTATATCCAAGTGCCATTCTCCCCAACC 81214

Qy     330 tccaggagtgagagtttctttagtatgtcttggatcctatgcccactttgtgaccagggc 389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81215 TCCAGGAGTGAGAGTTTCTTTAGTATGTCTTGGATCCTATGCCCACTTTGTGACCAGGGC 81274

Qy     390 aagaggg 396
           |||||||
Db   81275 AAGAGGG 81281


RESULT   2
AP000848/c
LOCUS       AP000848              186962 bp    DNA     linear   PRI 21-JUN-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-775D16,
            complete sequence.
ACCESSION   AP000848
VERSION     AP000848.4  GI:14517575
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:RP11-775D16.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 186962)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jun 20, 2001 this sequence version replaced gi:11136473.
FEATURES             Location/Qualifiers
     source          1. .186962
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-775D16"
BASE COUNT    57359 a  37204 c  36384 g  56015 t
ORIGIN


  Query Match             91.5%;  Score 362.2;  DB 9;  Length 186962;
  Best Local Similarity   99.2%;  Pred. No. 1.1e-98;
  Matches  364;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      30 cgaggtacatctgaattacaatagatgtgctaagatggtatgtttcggtgtgtattcttt 89
           ||  |||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db  124992 CGTAGTACATCTGAATTACAATAGATGTGCTAAGATGGTATGTGTCGGTGTGTATTCTTT 124933

Qy      90 ggttgcaagcaccaggcatgagttttggtgaatttaatcaagagagtaattcgttacaag 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124932 GGTTGCAAGCACCAGGCATGAGTTTTGGTGAATTTAATCAAGAGAGTAATTCGTTACAAG 124873

Qy     150 aacttggggtttctgtgcaacccaagaaaagtctgaaaataaccaaatttagaaaatgta 209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124872 AACTTGGGGTTTCTGTGCAACCCAAGAAAAGTCTGAAAATAACCAAATTTAGAAAATGTA 124813

Qy     210 aggaatcaggggggtttcaagactctagatagcaagaagtggaaaatcctgtctccaggga 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124812 AGGAATCAGGGGGTTTCAAGACTCTAGATAGCAAGAAGTGGAAAATCCTGTCTCCAGGGA 124753

Qy     270 acctcctcaggggagaatcagctccagctggtttatatccaagtgccattctccccaacc 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124752 ACCTCCTCAGGGGAGAATCAGCTCCAGCTGGTTTATATCCAAGTGCCATTCTCCCCAACC 124693

Qy     330 tccaggagtgagagtttctttagtatgtcttggatcctatgcccactttgtgaccagggc 389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124692 TCCAGGAGTGAGAGTTTCTTTAGTATGTCTTGGATCCTATGCCCACTTTGTGACCAGGGC 124633

Qy     390 aagaggg 396
           |||||||
Db  124632 AAGAGGG 124626


RESULT   3
AP000663
LOCUS       AP000663              115762 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-3F20 map 11q21, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AP000663
VERSION     AP000663.2  GI:8118852
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:CMB9-3F20.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 115762)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 115,762 genomic DNA of 11q21
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 115762)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997537.
```

```
------------- Genome Center
    Center: RIKEN Genomic Sciences Center(GSC)
    Center code: RIKEN
    Web site: http://hgp.gsc.riken.go.jp/
    Contact: hattori@gsc.riken.go.jp
------------- Project Information
    Center project name: HumDraft11
    Center clone name: CMB9-3F20
------------- Summary Statistics
    Sequencing vector: PCR products; 100% of reads
    Chemistry: Dye-terminator ET-amersham; 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 105451 bases at least Q40
    Consensus quality: 110431 bases at least Q30
    Consensus quality: 112519 bases at least Q20
    Insert size: 113762; sum-of-contigs
    Quality coverage: 4.53x in Q20 bases; sum-of-contigs
-------------.
NOTE: This is a 'working draft' sequence. It currently consists of
21 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved.
         1     27996 contig of  27996 bp in length
     28097     44759 contig of  16663 bp in length
     44860     53606 contig of   8747 bp in length
     53707     61270 contig of   7564 bp in length
     61371     66307 contig of   4937 bp in length
     66408     71338 contig of   4931 bp in length
     71439     77284 contig of   5846 bp in length
     77385     81759 contig of   4375 bp in length
     81860     86481 contig of   4622 bp in length
     86582     89025 contig of   2444 bp in length
     89126     92534 contig of   3409 bp in length
     92635     95988 contig of   3354 bp in length
     96089     99742 contig of   3654 bp in length
     99843    103044 contig of   3202 bp in length
    103145    105701 contig of   2557 bp in length
    105802    108757 contig of   2956 bp in length
    108858    110984 contig of   2127 bp in length
    111085    112926 contig of   1842 bp in length
    113027    114410 contig of   1384 bp in length
    114511    114563 contig of     53 bp in length
    114664    115762 contig of   1099 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1    27996: contig of 27996 bp in length
*     27997 28096: gap of      100 bp
*     28097    44759: contig of 16663 bp in length
*     44760 44859: gap of      100 bp
*     44860    53606: contig of 8747 bp in length
*     53607 53706: gap of      100 bp
*     53707    61270: contig of 7564 bp in length
*     61271 61370: gap of      100 bp
*     61371    66307: contig of 4937 bp in length
*     66308 66407: gap of      100 bp
*     66408    71338: contig of 4931 bp in length
*     71339 71438: gap of      100 bp
*     71439    77284: contig of 5846 bp in length
*     77285 77384: gap of      100 bp
*     77385    81759: contig of 4375 bp in length
*     81760 81859: gap of      100 bp
*     81860    86481: contig of 4622 bp in length
*     86482 86581: gap of      100 bp
*     86582    89025: contig of 2444 bp in length
*     89026 89125: gap of      100 bp
*     89126    92534: contig of 3409 bp in length
*     92535 92634: gap of      100 bp
*     92635    95988: contig of 3354 bp in length
*     95989 96088: gap of      100 bp
*     96089    99742: contig of 3654 bp in length
*     99743 99842: gap of      100 bp
*     99843   103044: contig of 3202 bp in length
*    103045 103144: gap of      100 bp
*    103145   105701: contig of 2557 bp in length
*    105702 105801: gap of      100 bp
*    105802   108757: contig of 2956 bp in length
*    108758 108857: gap of      100 bp
*    108858   110984: contig of 2127 bp in length
*    110985 111084: gap of      100 bp
*    111085   112926: contig of 1842 bp in length
*    112927 113026: gap of      100 bp
*    113027   114410: contig of 1384 bp in length
*    114411 114510: gap of      100 bp
*    114511   114563: contig of 53 bp in length
*    114564 114663: gap of      100 bp
*    114664   115762: contig of 1099 bp in length.
FEATURES             Location/Qualifiers
     source          1. .115762
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q21"
                     /clone="CMB9-3F20"
     misc_feature    1. .27996
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    28097. .44759
                     /note="assembly_fragment"
     misc_feature    44860. .53606
                     /note="assembly_fragment"
     misc_feature    53707. .61270
                     /note="assembly_fragment"
     misc_feature    61371. .66307
                     /note="assembly_fragment"
     misc_feature    66408. .71338
                     /note="assembly_fragment"
     misc_feature    71439. .77284
                     /note="assembly_fragment"
     misc_feature    77385. .81759
                     /note="assembly_fragment"
     misc_feature    81860. .86481
                     /note="assembly_fragment"
     misc_feature    86582. .89025
                     /note="assembly_fragment"
     misc_feature    89126. .92534
                     /note="assembly_fragment"
     misc_feature    92635. .95988
                     /note="assembly_fragment"
     misc_feature    96089. .99742
                     /note="assembly_fragment"
     misc_feature    99843. .103044
                     /note="assembly_fragment"
     misc_feature    103145. .105701
                     /note="assembly_fragment"
     misc_feature    105802. .108757
                     /note="assembly_fragment"
     misc_feature    108858. .110984
                     /note="assembly_fragment"
     misc_feature    111085. .112926
                     /note="assembly_fragment"
     misc_feature    113027. .114410
                     /note="assembly_fragment"
     misc_feature    114511. .114563
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    114664. .115762
                     /note="assembly_fragment"
BASE COUNT    34122 a  22523 c  22320 g  34797 t   2000 others
```

```
ORIGIN


  Query Match             91.1%;  Score 360.6;  DB 2;  Length 115762;
  Best Local Similarity   98.9%;  Pred. No. 3.2e-98;
  Matches  363;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      30 cgaggtacatctgaattacaatagatgtgctaagatggtatgtttcggtgtgtattcttt 89
           ||  |||||||||||||||||||||||||||||||||||| | ||||||||||||||
Db   68526 CGTAGTACATCTGAATTACAATAGATGTGCTAAGATGGTATGTGTTGGTGTGTATTCTTT 68585

Qy      90 ggttgcaagcaccaggcatgagttttggtgaatttaatcaagagagtaattcgttacaag 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68586 GGTTGCAAGCACCAGGCATGAGTTTTGGTGAATTTAATCAAGAGAGTAATTCGTTACAAG 68645

Qy     150 aacttggggtttctgtgcaacccaagaaaagtctgaaaataaccaaatttagaaaatgta 209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68646 AACTTGGGGTTTCTGTGCAACCCAAGAAAAGTCTGAAAATAACCAAATTTAGAAAATGTA 68705

Qy     210 aggaatcaggggggtttcaagactctagatagcaagaagtggaaaatcctgtctccaggga 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68706 AGGAATCAGGGGGTTTCAAGACTCTAGATAGCAAGAAGTGGAAAATCCTGTCTCCAGGGA 68765

Qy     270 acctcctcaggggagaatcagctccagctggtttatatccaagtgccattctccccaacc 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68766 ACCTCCTCAGGGGAGAATCAGCTCCAGCTGGTTTATATCCAAGTGCCATTCTCCCCAACC 68825

Qy     330 tccaggagtgagagtttctttagtatgtcttggatcctatgcccactttgtgaccagggc 389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68826 TCCAGGAGTGAGAGTTTCTTTAGTATGTCTTGGATCCTATGCCCACTTTGTGACCAGGGC 68885

Qy     390 aagaggg 396
           |||||||
Db   68886 AAGAGGG 68892


RESULT   4
AP001320
LOCUS       AP001320              172830 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-799O4 map 11q14, WORKING
            DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION   AP001320
VERSION     AP001320.2  GI:8117247
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-799O4.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172830)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 172,830 genomic DNA of 11q14
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 172830)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7209763.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: RP11-799O4
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 145836 bases at least Q40
                Consensus quality: 159404 bases at least Q30
                Consensus quality: 166388 bases at least Q20
                Insert size: 169730; sum-of-contigs
                Quality coverage: 4.17x in Q20 bases; sum-of-contigs
            --------------.
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
         1     15732 contig of  15732 bp in length
     15833     34101 contig of  18269 bp in length
     34202     47479 contig of  13278 bp in length
     47580     57756 contig of  10177 bp in length
     57857     69426 contig of  11570 bp in length
     69527     80185 contig of  10659 bp in length
     80286     88989 contig of   8704 bp in length
     89090     99362 contig of  10273 bp in length
     99463    103433 contig of   3971 bp in length
    103534    109064 contig of   5531 bp in length
    109165    113855 contig of   4691 bp in length
    113956    119140 contig of   5185 bp in length
    119241    124661 contig of   5421 bp in length
    124762    126822 contig of   2061 bp in length
    126923    131337 contig of   4415 bp in length
    131438    135078 contig of   3641 bp in length
    135179    138134 contig of   2956 bp in length
    138235    142123 contig of   3889 bp in length
    142224    144719 contig of   2496 bp in length
    144820    149030 contig of   4211 bp in length
    149131    151134 contig of   2004 bp in length
    151235    153751 contig of   2517 bp in length
    153852    156504 contig of   2653 bp in length
    156605    159583 contig of   2979 bp in length
    159684    161669 contig of   1986 bp in length
    161770    164135 contig of   2366 bp in length
    164236    165854 contig of   1619 bp in length
    165955    167495 contig of   1541 bp in length
    167596    168771 contig of   1176 bp in length
    168872    170322 contig of   1451 bp in length
    170423    171562 contig of   1140 bp in length
    171663    172830 contig of   1168 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1     15732: contig of 15732 bp in length
*     15733 15832: gap of      100 bp
*     15833     34101: contig of 18269 bp in length
*     34102 34201: gap of      100 bp
*     34202     47479: contig of 13278 bp in length
*     47480 47579: gap of      100 bp
*     47580     57756: contig of 10177 bp in length
*     57757 57856: gap of      100 bp
*     57857     69426: contig of 11570 bp in length
*     69427 69526: gap of      100 bp
*     69527     80185: contig of 10659 bp in length
*     80186 80285: gap of      100 bp
*     80286     88989: contig of 8704 bp in length
*     88990 89089: gap of      100 bp
*     89090     99362: contig of 10273 bp in length
*     99363 99462: gap of      100 bp
```

```
          *    99463   103433: contig of 3971 bp in length
          *   103434 103533: gap of      100 bp
          *   103534   109064: contig of 5531 bp in length
          *   109065 109164: gap of      100 bp
          *   109165   113855: contig of 4691 bp in length
          *   113856 113955: gap of      100 bp
          *   113956   119140: contig of 5185 bp in length
          *   119141 119240: gap of      100 bp
          *   119241   124661: contig of 5421 bp in length
          *   124662 124761: gap of      100 bp
          *   124762   126822: contig of 2061 bp in length
          *   126823 126922: gap of      100 bp
          *   126923   131337: contig of 4415 bp in length
          *   131338 131437: gap of      100 bp
          *   131438   135078: contig of 3641 bp in length
          *   135079 135178: gap of      100 bp
          *   135179   138134: contig of 2956 bp in length
          *   138135 138234: gap of      100 bp
          *   138235   142123: contig of 3889 bp in length
          *   142124 142223: gap of      100 bp
          *   142224   144719: contig of 2496 bp in length
          *   144720 144819: gap of      100 bp
          *   144820   149030: contig of 4211 bp in length
          *   149031 149130: gap of      100 bp
          *   149131   151134: contig of 2004 bp in length
          *   151135 151234: gap of      100 bp
          *   151235   153751: contig of 2517 bp in length
          *   153752 153851: gap of      100 bp
          *   153852   156504: contig of 2653 bp in length
          *   156505 156604: gap of      100 bp
          *   156605   159583: contig of 2979 bp in length
          *   159584 159683: gap of      100 bp
          *   159684   161669: contig of 1986 bp in length
          *   161670 161769: gap of      100 bp
          *   161770   164135: contig of 2366 bp in length
          *   164136 164235: gap of      100 bp
          *   164236   165854: contig of 1619 bp in length
          *   165855 165954: gap of      100 bp
          *   165955   167495: contig of 1541 bp in length
          *   167496 167595: gap of      100 bp
          *   167596   168771: contig of 1176 bp in length
          *   168772 168871: gap of      100 bp
          *   168872   170322: contig of 1451 bp in length
          *   170323 170422: gap of      100 bp
          *   170423   171562: contig of 1140 bp in length
          *   171563 171662: gap of      100 bp
          *   171663   172830: contig of 1168 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172830
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q14"
                     /clone="RP11-799O4"
     misc_feature    1. .15732
                     /note="assembly_fragment"
     misc_feature    15833. .34101
                     /note="assembly_fragment"
     misc_feature    34202. .47479
                     /note="assembly_fragment"
     misc_feature    47580. .57756
                     /note="assembly_fragment"
     misc_feature    57857. .69426
                     /note="assembly_fragment"
     misc_feature    69527. .80185
                     /note="assembly_fragment"
     misc_feature    80286. .88989
                     /note="assembly_fragment"
     misc_feature    89090. .99362
                     /note="assembly_fragment"
     misc_feature    99463. .103433
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    103534. .109064
                     /note="assembly_fragment"
     misc_feature    109165. .113855
                     /note="assembly_fragment"
     misc_feature    113956. .119140
                     /note="assembly_fragment"
     misc_feature    119241. .124661
                     /note="assembly_fragment"
     misc_feature    124762. .126822
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    126923. .131337
                     /note="assembly_fragment"
     misc_feature    131438. .135078
                     /note="assembly_fragment"
     misc_feature    135179. .138134
                     /note="assembly_fragment"
     misc_feature    138235. .142123
                     /note="assembly_fragment"
     misc_feature    142224. .144719
                     /note="assembly_fragment"

  Query Match             88.0%;  Score 348.6;  DB 2;  Length 172830;
  Best Local Similarity   98.6%;  Pred. No. 1.4e-94;
  Matches  362;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy      30 cgaggtacatctgaattacaatagatgtgctaagatggtatgtttcggtgtgtattcttt 89
           ||  |||||||||||||||||||||||||||||||||||| | ||||||||||||||
Db   56948 CGTAGTACATCTGAATTACAATAGATGTGCTAAGATGGTATGTGTTGGTGTGTATTCTTT 57007

Qy      90 ggttgcaagcaccaggcatgagttttggtgaatttaatcaagagagtaattcgttacaag 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57008 GGTTGCAAGCACCAGGCATGAGTTTTGGTGAATTTAATCAAGAGAGTAATTCGTTACAAG 57067

Qy     150 aacttggggtttctgtgcaacccaagaaaagtctgaaaataaccaaatttagaaaatgta 209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57068 AACTTGGGGTTTCTGTGCAACCCAAGAAAAGTCTGAAAATAACCAAATTTAGAAAATGTA 57127

Qy     210 aggaatcagggggtttcaagactctagatagcaagaagtggaaaatcctgtctccaggga 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57128 AGGAATCAGGGGGTTTCAAGACTCTAGATAGCAAGAAGTGGAAAATCCTGTCTCCAGGG- 57186

Qy     270 acctcctcaggggagaatcagctccagctggtttatatccaagtgccattctccccaacc 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57187 ACCTCCTCAGGGGAGAATCAGCTCCAGCTGGTTTATATCCAAGTGCCATTCTCCCCAACC 57246

Qy     330 tccaggagtgagagtttctttagtatgtcttggatcctatgcccactttgtgaccagggc 389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57247 TCCAGGAGTGAGAGTTTCTTTAGTATGTCTTGGATCCTATGCCCACTTTGTGACCAGGGC 57306

Qy     390 aagaggg 396
           |||||||
Db   57307 AAGAGGG 57313


RESULT   5
AC017280/c
LOCUS       AC017280               45029 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017280
VERSION     AC017280.1  GI:6553706
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 45029)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210143 by the submitter.
            For more information on this record e-mail to fly@celera.com.
```

```
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .45029
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    12625 a   9495 c   9780 g  13129 t
ORIGIN


  Query Match           11.5%;  Score 45.4;  DB 2;  Length 45029;
  Best Local Similarity   51.8%;  Pred. No. 0.0084;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      49 aatagatgtgctaagatggtatgtttcggtgtgtattctttggttgcaagcaccaggcat 108
           || ||| |  ||   ||  |||||   |  || | ||  |||     |||||    | ||
Db   18059 AACAGAAGGCCTTTAATTATATGTGCAGTAGTTTGTTTGTTGAGCTTAAGCAAGCCGTAT 18000

Qy     109 gagttttggtgaatttaatcaagagagtaattcgttacaagaacttggggtttctgtgca 168
            | | |   |  |  |||| || |    |   |   | | |    ||| ||   |   |
Db   17999 AAATATAAATTCAGCTAATGAATAACAAATAGCCGAAAATGCTAATGGAGTAGATCAAAA 17940

Qy     169 acccaagaaaagtctgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaa 228
           | | || |||||  | |||  ||  | | ||| || | ||||  ||| | ||  | |  |
Db   17939 ATCAAAAAAAAGAATAAAACAAATTATAATTATAATAAGTAATAAATAAAGGAATATGTA 17880

Qy     229 gactctagatagcaagaag 247
           |  | |||| || |  ||
Db   17879 TATACCAGATGGCGACGAG 17861


RESULT   6
AC006168
LOCUS       AC006168               68626 bp    DNA     linear   HTG 10-FEB-2000
DEFINITION  Drosophila melanogaster chromosome X clone BACR48C06 (D485) RPCI-98
            48.C.6 map 3C-3C strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 55 unordered pieces.
ACCESSION   AC006168
VERSION     AC006168.3  GI:6957989
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 68626)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68626)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,
            Zieran,L.L. and Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 11, 2000 this sequence version replaced gi:5912635.
            For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1      730: contig of 730 bp in length
*      731      810: gap of unknown length
*      811     1566: contig of 756 bp in length
*     1567     1646: gap of unknown length
*     1647     2491: contig of 845 bp in length
*     2492     2571: gap of unknown length
*     2572     3433: contig of 862 bp in length
*     3434     3513: gap of unknown length
*     3514     4268: contig of 755 bp in length
*     4269     4348: gap of unknown length
*     4349     4905: contig of 557 bp in length
*     4906     4985: gap of unknown length
*     4986     5654: contig of 669 bp in length
*     5655     5734: gap of unknown length
*     5735     6639: contig of 905 bp in length
*     6640     6719: gap of unknown length
*     6720     7309: contig of 590 bp in length
*     7310     7389: gap of unknown length
*     7390     8144: contig of 755 bp in length
*     8145     8224: gap of unknown length
*     8225     9026: contig of 802 bp in length
*     9027     9106: gap of unknown length
*     9107    10073: contig of 967 bp in length
*    10074    10153: gap of unknown length
*    10154    11779: contig of 1626 bp in length
*    11780    11859: gap of unknown length
*    11860    12725: contig of 866 bp in length
*    12726    12805: gap of unknown length
*    12806    14383: contig of 1578 bp in length
*    14384    14463: gap of unknown length
*    14464    15848: contig of 1385 bp in length
*    15849    15928: gap of unknown length
*    15929    16752: contig of 824 bp in length
*    16753    16832: gap of unknown length
*    16833    17476: contig of 644 bp in length
*    17477    17556: gap of unknown length
*    17557    18580: contig of 1024 bp in length
*    18581    18660: gap of unknown length
*    18661    20474: contig of 1814 bp in length
*    20475    20554: gap of unknown length
*    20555    22413: contig of 1859 bp in length
*    22414    22493: gap of unknown length
*    22494    24150: contig of 1657 bp in length
*    24151    24230: gap of unknown length
*    24231    25965: contig of 1735 bp in length
*    25966    26045: gap of unknown length
*    26046    27842: contig of 1797 bp in length
*    27843    27922: gap of unknown length
*    27923    29292: contig of 1370 bp in length
*    29293    29372: gap of unknown length
*    29373    31590: contig of 2218 bp in length
*    31591    31670: gap of unknown length
*    31671    33808: contig of 2138 bp in length
*    33809    33888: gap of unknown length
*    33889    35388: contig of 1500 bp in length
*    35389    35468: gap of unknown length
*    35469    37677: contig of 2209 bp in length
*    37678    37757: gap of unknown length
*    37758    39729: contig of 1972 bp in length
*    39730    39809: gap of unknown length
*    39810    42207: contig of 2398 bp in length
```

```
            *    42208    42287: gap of unknown length
            *    42288    45740: contig of 3453 bp in length
            *    45741    45820: gap of unknown length
            *    45821    52324: contig of 6504 bp in length
            *    52325    52404: gap of unknown length
            *    52405    53166: contig of 762 bp in length
            *    53167    53246: gap of unknown length
            *    53247    53978: contig of 732 bp in length
            *    53979    54058: gap of unknown length
            *    54059    54639: contig of 581 bp in length
            *    54640    54719: gap of unknown length
            *    54720    55377: contig of 658 bp in length
            *    55378    55457: gap of unknown length
            *    55458    56074: contig of 617 bp in length
            *    56075    56154: gap of unknown length
            *    56155    56785: contig of 631 bp in length
            *    56786    56865: gap of unknown length
            *    56866    57505: contig of 640 bp in length
            *    57506    57585: gap of unknown length
            *    57586    58418: contig of 833 bp in length
            *    58419    58498: gap of unknown length
            *    58499    59104: contig of 606 bp in length
            *    59105    59184: gap of unknown length
            *    59185    59825: contig of 641 bp in length
            *    59826    59905: gap of unknown length
            *    59906    60665: contig of 760 bp in length
            *    60666    60745: gap of unknown length
            *    60746    61377: contig of 632 bp in length
            *    61378    61457: gap of unknown length
            *    61458    62118: contig of 661 bp in length
            *    62119    62198: gap of unknown length
            *    62199    62791: contig of 593 bp in length
            *    62792    62871: gap of unknown length
            *    62872    63524: contig of 653 bp in length
            *    63525    63604: gap of unknown length
            *    63605    64223: contig of 619 bp in length
            *    64224    64303: gap of unknown length
            *    64304    64965: contig of 662 bp in length
            *    64966    65045: gap of unknown length
            *    65046    65693: contig of 648 bp in length
            *    65694    65773: gap of unknown length
            *    65774    66389: contig of 616 bp in length
            *    66390    66469: gap of unknown length
            *    66470    67047: contig of 578 bp in length
            *    67048    67127: gap of unknown length
            *    67128    67841: contig of 714 bp in length
            *    67842    67921: gap of unknown length
            *    67922    68626: contig of 705 bp in length.
FEATURES             Location/Qualifiers
     source          1. .68626
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="3C-3C"
                     /clone="BACR48C06 (D485) RPCI-98 48.C.6"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6"
BASE COUNT    17635 a  14487 c  13985 g  18195 t   4324 others
ORIGIN

  Query Match           11.5%;  Score 45.4;  DB 2;  Length 68626;
  Best Local Similarity   51.8%;  Pred. No. 0.0085;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      49 aatagatgtgctaagatggtatgtttcggtgtgtattctttggttgcaagcaccaggcat 108
           || ||| |  ||   ||  |||||  |  || | ||  |||     |||||    | ||
Db   27223 AACAGAAGGCCTTTAATTATATGTGCAGTAGTTTGTTTGTTGAGCTTAAGCAAGCCGTAT 27282

Qy     109 gagttttggtgaatttaatcaagagagtaattcgttacaagaacttggggtttctgtgca 168
           | | |   |  |  |||| || |   |   |  | | |    ||| ||   |    |
```

```
Db   27283 AAATATAAATTCAGCTAATGAATAACAAATAGCCGAAAATGCTAATGGAGTAGATCAAAA 27342

Qy     169 acccaagaaaagtctgaaaataaccaaatttagaaaatgtaaggaatcagggggtttcaa 228
           | | || |||||  | |||  ||  | | ||| || | ||||  ||| | ||  | | |
Db   27343 ATCAAAAAAAAGAATAAAACAAATTATAATTATAATAAGTAATAAATAAAGGAATATGTA 27402

Qy     229 gactctagatagcaagaag 247
           |  | ||||  || |  ||
Db   27403 TATACCAGATGGCGACGAG 27421


RESULT   7
DMBR43E12/c
LOCUS       DMBR43E12              102621 bp    DNA     linear   INV 07-FEB-2000
DEFINITION  Drosophila melanogaster BAC clone BACR43E12.
ACCESSION   AL138971
VERSION     AL138971.1  GI:6940335
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 102621)
  AUTHORS   Ferraz,C., Vidal,S., Brun,C., Bucheton,A. and Demaille,J.G.
  TITLE     Sequencing the distal X chromosome of Drosophila melanogaster
  JOURNAL   Unpublished
  REMARK    Institut de Genetique Humaine -UPR 1142- CNRS, 141,rue de la
            Cardonille, 34396 Montpellier Cedex 5, France
REFERENCE   2  (bases 1 to 102621)
  AUTHORS   Benos,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2000) European Drosophila Genome Sequencing
            Consortium
COMMENT     Sequence submitted by Takis Benos, EMBL Outstation - The EBI,
            Hinxton, Cambridge, CB10 1SD, U.K.
            E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome
            Sequencing Consortium. For further information see the European
            Drosophila Genome Sequencing Consortium's web site:
            http://edgp.ebi.ac.uk/.
            The syntax for the representation of annotation used in this record
            is documented at:
            ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_annotation.README
            Coding sequences are predicted from computer analysis, using both
            gene and CDS prediction programs and matches to other sequences.
            These predictions and matches have been evaluated by the annotators
            and may have been refined by hand (in which case a Genefinder
            prediction will have no score. The annotators have also used their
            judgement on what matches to represent in this record. A far more
            complete annotation record is available from FlyBase
            (http://flybase.bio.indiana.edu/) through the FlyBase Annotation
            Object linked by the db_xref qualifier in the Feature Table.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            clone BACR43E12.  It may be shorter, since we are minimising the
            overlap between clones to 100 bases, by trimming them. Clone
            BACR43E12 overlaps to the left with clone BACN33B1 The true left
            end of clone BACR43E12 is unknown
            Clone BACR43E12 overlaps to the right with clone 100G7 The true
            right end of clone BACR43E12 is unknown
            Sequence in absolute orientation with respect to chromosome
            Clone=BACR43E12; Contig ID=28; Length=102621; Status=Finished.
FEATURES             Location/Qualifiers
     source          1. .102621
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BAC BACR43E12"
     gene            6318. .7103
                     /gene="EG:BACR43E12.1"
     CDS             6318. .7103
                     /gene="EG:BACR43E12.1"
                     /note="/prediction=(method:''genefinder'',
                     version:''084'', score:''29.13'');
                     /prediction=(method:''genscan'', version:''1.0'',
```

```
                score:''84.99''); /match=(desc:''MYELOBLAST KIAA0252
                (FRAGMENT)'', species:''Homo sapiens (Human)'',
                ranges:(query:6441. .6860,
                target:SPTREMBL::Q92541:298. .437, score:''210.00'')),
                method:''blastx'', version:''1.4.9'')"
                /codon_start=1
                /protein_id="CAB72279.1"
                /db_xref="GI:6940336"
                /translation="MSNQRSRHVRRPAAEIVRELRNLRASPLRIDDENESWVNGYEKP
                PTPDLPVTSRDQLELLRLSRHRIGLLLVRPAFEQAVTGCFVRVNVSGQGELPDHRIAE
                VLGICELDFGYKVEQIPTNVALRLRYDDLEMQHEINDVSNLNFTQEEFELWRDNCVNQ
                AISPPTTHILTRKKVELYNALQLEAKPLSLIQRTFSFALRPQQKIGIMERHGVVYPWQ
                LNHSLPFVSQSPTENPSRGKADTVDEEEDPGAAAPLLPKTEPN"
repeat_unit     complement(20231. .28682)
                /gene="FBgn0000155;roo"
                /function="transposon"
gene            complement(20231. .28682)
                /gene="EG:BACR43E12.2"
gene            complement(41307. .48656)
                /gene="EG:BACR43E12.3"
repeat_unit     complement(41307. .48656)
                /gene="FBgn0002697;mgd1"
                /function="transposon"
gene            complement(68339. .69170)
                /gene="EG:BACR43E12.7"
CDS             complement(join(68339. .69036,69092. .69170))
                /gene="EG:BACR43E12.7"
                /note="/prediction=(method:''genefinder'',
                version:''084'', score:''34.12'');
                /prediction=(method:''genscan'', version:''1.0'',
                score:''98.21''); the EST embl|AA696429|AA696429 comes
                from the 3' UTR"
                /codon_start=1
                /protein_id="CAB72280.1"
                /db_xref="GI:6940337"
                /translation="MQILHNLFALLVVLLAILANCDARRSAYRLPKLQGRSASGSPAP
                QDVGSGRDFMKKFMLMRGLFQQPPSDNVIVITQPSTTTTTTTPTGSPTTTTSTTTTTT
                PTSNSTGRSLQESVHRQQPQDDDEAQETGVTTLTGRVLDDRLDLPDGEEEPQELVTHK
                HTQKPKKFNMKSVKLNNRKKNLVKKTHRRVHKHPQKKKRHQKRKLNKKQSAANQQRNS
                EPAVNFIEPYHQDRGQSQAGGQSTSDSRLQRIWRRFQIAN"
gene            complement(71879. .72871)
                /gene="EG:BACR43E12.6"
CDS             complement(71879. .72871)
                /gene="EG:BACR43E12.6"
                /note="/prediction=(method:''genefinder'',
                version:''084'', score:''38.91'');
                /prediction=(method:''genscan'', version:''1.0'',
                score:''99.03'')"
                /codon_start=1
                /protein_id="CAB72281.1"
                /db_xref="GI:6940338"
                /translation="MQPPPKTYNWLLLLLVLLGISLARTAPRQATSTYVIPMSWNDLI
                GTFNFPEPQEKSHEIPEPRAFSVSLDDMSPDGQQADDLQQVLMLTSGLTPAQIAQRSQ
                GRNNEGRFFGRGGTTVVVITGGGAANNTNNNSFVSTPASTTTTGTRTSEFIRTPIAYP
                AGLPMQYFRKRQDDGTSMSYPEFLRWNEAAFYGGDLGLSGSNGAYGLSGVPLVPITVG
                NEVRYVPMNLRMFRQLVGDPTTVASPPPPSIREQDDQLEDDDISALSLGPELEVLDES
                PNEGGQEVAGSGHSVADSPGFGILGQRLRQRPLVRRRPLQTLAQNIRRVQYVRK"
gene            complement(78912. .79974)
                /gene="EG:BACR43E12.5"
CDS             complement(join(78912. .79664,79735. .79811,79896. .79974))
                /gene="EG:BACR43E12.5"
                /note="/prediction=(method:''genefinder'',
                version:''084'', score:''49.15'');
                /prediction=(method:''genscan'', version:''1.0'')"
                /codon_start=1
                /protein_id="CAB72282.1"
                /db_xref="GI:6940339"
                /translation="MCGKGFGMILVLIWATALSSDLSLVSGQTVAPLTSSTSAPPKME
                VKPTKPMIVTGFITKSGNIYEIEDKRGSIGSIEGRQADQEQIVCNYGNVVIYSDVPCD
                QVKNVRVGEVKPLKDDRVEVTTEKNHADGEGEQQSQDMQQEQQQQDAAEDQDQEQDQD
                HQSVQSNHPRGQQNNRRRRRRPQQQQQQQRLQQQRRRRQQQQKLQQRRRNGNGNGNNS
                LRRRRNRNNNNNNLKRQQQRRRRPGNNNNNNRRRLNNNNSNNNNKQRQQQHRRQQQQQ
                QRRRRPNRNINRQQQQQRRRLHDDNI"
```

```
     gene            88303. .90223
                     /gene="EG:BACR43E12.4"
     CDS             join(88303. .88527,89561. .90223)
                     /gene="EG:BACR43E12.4"
                     /note="/match=(desc:''ALPHA-DYSTROBREVIN-3'',
                     species:''Homo sapiens (Human)'',
                     ranges:(query:89621. .89761,
                     target:SPTREMBL::O75697:241. .287, score:''105.00'')),
                     method:''blastx'', version:''1.4.9'');
                     /match=(desc:''REF(2)P PROTEIN'', species:''Drosophila
                     melanogaster (Fruit fly)'', ranges:(query:89612. .89713,
                     target:SWISS-PROT::P14199:122. .155, score:''107.00'')),
                     method:''blastx'', version:''1.4.9''); the EST
                     embl|AI062288|AI062288 comes from the 5' UTR"
                     /codon_start=1
                     /protein_id="CAB72283.1"
                     /db_xref="GI:6940340"
                     /translation="MHSTSFRCGQQIRAAHAGRANIIIIIAAISIFGRLKGSRCPAGK
                     TRKHCAFSPHIRTYIVPPGFPQCTPFALRLVSGSPSMPRRRQIGHTSGHCNVRCDGCG
                     NNRMTFYRYKCLHCLDYDLCSDCKENGVSNGLHSLDHPLQCLMDRDALELHFAGEPIP
                     ILCADSFTCPVCGEMGFSVEDLRTHCQDNHRMARTVCICPVCAAVPLSQPSHIAHIAN
                     HLMFSPSHRATGDPMIDISATSQAEGSSLPQNSAVSSGSGAQHTFSSSENSQVRFLLP
                     GNRAPLADSDDFNMDNWEV"
BASE COUNT    28899 a  20617 c  21237 g  31868 t
ORIGIN

  Query Match           11.5%;  Score 45.4;  DB 3;  Length 102621;
  Best Local Similarity  51.8%;  Pred. No. 0.0086;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      49 aatagatgtgctaagatggtatgtttcggtgtgtattctttggttgcaagcaccaggcat 108
           || ||| |  ||   ||  |||||   |  || | ||  |||     |||||    | ||
Db   83966 AACAGAAGGCCTTTAATTATATGTGCAGTAGTTTGTTTGTTGAGCTTAAGCAAGCCGTAT 83907

Qy     109 gagttttggtgaatttaatcaagagagtaattcgttacaagaacttggggtttctgtgca 168
           | | |   |  |  |||| || |    |   |   | | |    ||| ||   |    |
Db   83906 AAATATAAATTCAGCTAATGAATAACAAATAGCCGAAAATGCTAATGGAGTAGATCAAAA 83847

Qy     169 acccaagaaaagtctgaaaataaccaaatttagaaaaatgtaaggaatcaggggttttcaa 228
           | | || |||||  | |||  ||  | | ||| || | ||||  ||| | ||  | |  |
Db   83846 ATCAAAAAAAAGAATAAAACAAATTATAATTATAATAAGTAATAAATAAAGGAATATGTA 83787

Qy     229 gactctagatagcaagaag 247
           |  | |||| || |  ||
Db   83786 TATACCAGATGGCGACGAG 83768


RESULT   8
AC104148/c
LOCUS       AC104148              181443 bp    DNA     linear   INV 27-DEC-2001
DEFINITION  Drosophila melanogaster X BAC RP98-34O3 (Roswell Park Cancer
            Institute Drosophila BAC Library) complete sequence.
ACCESSION   AC104148
VERSION     AC104148.3  GI:17981668
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 181443)
  AUTHORS   Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
            Gocayne,J.D., Tabor,P., Williamson,A., Homsi,F.H.,
            Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
            Ayele,M.A., Scott,G.S., Worley,K.W., Amamatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
            Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
            Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
            Draper,H., Emery-Cohen,A., Ferriera,S., Garg,N.D.S., Houck,J.,
            Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
            Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
            Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
```

Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M., Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O., Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 181443)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 181443)
AUTHORS Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Submitted (27-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Dec 27, 2001 this sequence version replaced gi:17940811.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

```
FEATURES             Location/Qualifiers
     source          1. .181443
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /clone="RP98-34O3"
     misc_feature    64174
                     /function="low coverage"
     misc_feature    64176
                     /function="unsure base"
     misc_feature    145878. .145937
                     /function="unresolved tandem repeat"
BASE COUNT    51551 a  37954 c  38769 g  53169 t
ORIGIN
```

```
  Query Match            11.5%;  Score 45.4;  DB 3;  Length 181443;
  Best Local Similarity  51.8%;  Pred. No. 0.0087;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      49 aatagatgtgctaagatggtatgtttcggtgtgtattctttggttgcaagcaccaggcat 108
           || ||| |  ||   ||  |||||   |  || | ||  |||     |||||    | ||
Db 101132 AACAGAAGGCCTTTAATTATATGTGCAGTAGTTTGTTTGTTGAGCTTAAGCAAGCCGTAT 101073

Qy     109 gagttttggtgaatttaatcaagagagtaattcgttacaagaacttggggtttctgtgca 168
           | | |   |  |  |||| || |    |   |   | | |    ||| ||   |    |
Db 101072 AAATATAAATTCAGCTAATGAATAACAAATAGCCGAAAATGCTAATGGAGTAGATCAAAA 101013

Qy     169 acccaagaaaagtctgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaa 228
           | | || |||||  | |||  ||  | | ||| || | ||||  ||| | ||  | |  |
Db 101012 ATCAAAAAAAAGAATAAAACAAATTATAATTATAATAAGTAATAAATAAAGGAATATGTA 100953

Qy     229 gactctagatagcaagaag 247
           |  | |||| || |  ||
Db 100952 TATACCAGATGGCGACGAG 100934
```

```
RESULT   9
AE003425/c
LOCUS       AE003425              300205 bp    DNA     linear   INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386054 section 9
            of 35, complete sequence.
ACCESSION   AE003425 AE002566
VERSION     AE003425.1  GI:7290341
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 300205)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
```

```
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 300205)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .300205
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
     mRNA            join(<12310. .12326,13904. .14602,14696. .15533,24623. .24826,
                     24897. .24980,25844. .25954,26032. .26269,26353. .26546,
                     27639. .27828,27900. .28343,29322. .29779)
                     /gene="sgg"
                     /product="CT8875"
                     /db_xref="FLYBASE:FBan0002621"
                     /db_xref="FLYBASE:FBgn0003371"
     gene            <12310. .>29779
                     /gene="sgg"
                     /note="CG2621"
                     /map="3B1-3B4"
                     /db_xref="FLYBASE:FBan0002621"
                     /db_xref="FLYBASE:FBgn0003371"
     CDS             join(12310. .12326,13904. .14602,14696. .15533,24623. .24826,
                     24897. .24980,25844. .25954,26032. .26269,26353. .26546,
                     27639. .27828,27900. .28343,29322. .29329)
                     /gene="sgg"
                     /note="sgg gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0002621"
                     /db_xref="FLYBASE:FBgn0003371"
                     /protein_id="AAF45801.1"
                     /db_xref="GI:7290342"
                     /translation="MFTFYTNINNTLINNNNNNNNTSNSNNNNNNVISQPIKIPLTER
                     FSSQTSTGSADSGVIVSSASQQQLQLPPPRSSSGSLSLPQAPPGGKWRQKQQRQQLLL
                     SQDSGIENGVTTRPSKAKDNQGAGKASHNATSSKESGAQSNSSSESLGSNCSEAQEQQ
                     RVRASSALELSSVDTPVIVGGVVSGGNSILRSRIKYKSTNSTGTQGFDVEDRIDEVDI
                     CDDDDVDCDDRGSEIEEEEEDQTEQEEEVDEVDAKPKNRLLPPDQAELTVAAAMARRR
                     DAKSLATDGHIYFPLLKISEDPHIDSKLINRKDGLQDTMYYLDEFGSPKLREKFARKQ
                     KQLLAKQQKQLMKRERRSEEQRKKRNTTVASNLAASGAVVDDTKDDYKQQPHCDTSSR
                     SKNNSVPNPPSSHLHQNHNHLVVDVQEDVDDVNVVATSDVDSGVVKMRRHSHDNHYDR
                     IPRSNAATITTRPQIDQQSSHHQNTEDVEQGAEPQIDGEADLDADADADSDGSGENVK
                     TAKLARTQSCKNQTGRDGSKITTVVATPGQGTDRVQEVSYTDTKVIGNGSFGVVFQAK
                     LCDTGELVAIKKVLQDRRFKNRELQIMRKLEHCNIVKLLYFFYSSGEKRDEVFLNLVL
                     EYIPETVYKVARQYAKTKQTIPINFIRLYMYQLFRSLAYIHSLGICHRDIKPQNLLLD
                     PETAVLKLCDFGSAKQLLHGEPNVSYICSRYYRAPELIFGAINYTTKIDVWSAGCVLA
                     ELLLGQPIFPGDSGVDQLVEVIKVLGTPTREQIREMNPNYTEFKFPQIKSHPWQKVFR
                     IRTPTEAINLVSLLLEYTPSARITPLKACAHPFFDELRMEGNHTLPNGRDMPPLFNFT
                     EHELSIQPSLVPQLLPKHLQNASGPGGNRPSAGGAASIAASGSTSVSSTGSGASVEGS
                     AQPQSQGTAAAAGSGSGGATAGTGGASAGGPGSGNNSSSGGASGAPSAVAAGGANAAV
                     AGGAGGGGGAGAATAAATATGAIGATNAGGANVTDS"
     mRNA            complement(join(35010. .35971,36069. .36458))
                     /gene="HLH3B"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product="CT8939"
                     /db_xref="FLYBASE:FBan0002655"
                     /db_xref="FLYBASE:FBgn0011276"
     gene            complement(<35010. .>36458)
                     /gene="HLH3B"
                     /note="CG2655"
                     /map="3B3-3B4"
                     /db_xref="FLYBASE:FBan0002655"
                     /db_xref="FLYBASE:FBgn0011276"
     CDS             complement(join(35166. .35971,36069. .36393))
                     /gene="HLH3B"
                     /note="HLH3B gene product; Nucleotide sequence of the
                     Celera sequence differs from the published sequence for
                     this transcript"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0002655"
                     /db_xref="FLYBASE:FBgn0011276"
                     /protein_id="AAF45802.1"
                     /db_xref="GI:7290343"
                     /translation="MAWLPSSSNGADNADERSTASSGSSGHSQTNESPRSGHLNGNGS
                     MRDTAAGRHPPALLRHATPHLQPKTESISDGDGDAELSDFSLNDTEEDEEDLRDYIVL
                     NGNQADANRSLSSSPRSHSRNGLLTAPASSGSSVGGSGGGGGNGSGGNASSGGGSGVG
                     ATGGVRKVFTNTRERWRQQNVSGAFAELRKLVPTHPPDKKLSKNEILRSAIKYIKLLT
                     GILEWQQRQAPSHPIRAQMEPNNNDNRMANGHAADGENLENPDVPPVRHIKCERTDGQ
                     MHRNGIGHGHANGNAGNDLLMIAPGAVVKSELLLESTLPLGHPLNGPPLPLTTAPLAM
                     AETQRISGTVSGVKSASGRSSKRRLKPEGGATDLSLGKRRRT"
     mRNA            complement(36852. .>37664)
                     /gene="EG:155E2.5"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product="CT8935"
                     /db_xref="FLYBASE:FBan0002652"
                     /db_xref="FLYBASE:FBgn0025838"
     gene            complement(<36852. .>37664)
                     /gene="EG:155E2.5"
                     /note="CG2652"
                     /map="3B3-3B4"
                     /db_xref="FLYBASE:FBan0002652"
                     /db_xref="FLYBASE:FBgn0025838"
     CDS             complement(36852. .37559)
                     /gene="EG:155E2.5"
                     /note="EG:155E2.5 gene product; Nucleotide sequence of the
                     Celera sequence differs from the published sequence for
                     this transcript"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0002652"
                     /db_xref="FLYBASE:FBgn0025838"
                     /protein_id="AAF45803.1"
                     /db_xref="GI:7290344"
                     /translation="MSYTNLLRHQNVVDDCQRVSCNRTPPITVTGRCVISRDIVIGCR
                     MEQCDPDMPYMLEPTWGVYLRPGRDGGDLSRFVRRVTFKMSPRLPLRLHVADSAPFEI
```

```
                GEVLGSDFPVEVQVQYMDARMSATSYIFRPRVVREGHAGICEEMLDKMIFVNPSPMMR
                QNLTPVLVPSANGAPGDRTSPQFAMESAMPETPGERKEQDRDREQVGDVARPSQPPAK
                QPKKRLSVGIPHPIGHQ"
     mRNA       join(42932. .43074,43135. .44174,44239. .44595,44658. .46372,
                46443. .46679,46744. .46884,46943. .>47014)
                /gene="per"
                /note="Nucleotide sequence of the Celera sequence differs
                from the published sequence for this transcript."
                /product="CT8963"
                /db_xref="FLYBASE:FBan0002647"
                /db_xref="FLYBASE:FBgn0003068"
     gene       <42932. .>47014
                /gene="per"
                /note="CG2647"
                /map="3B3-3B5"
                /db_xref="FLYBASE:FBan0002647"
                /db_xref="FLYBASE:FBgn0003068"
     CDS        join(42980. .43074,43135. .44174,44239. .44595,44658. .46372,
                46443. .46679,46744. .46884,46943. .47014)
                /gene="per"
                /note="per gene product; Nucleotide sequence of the Celera
                sequence differs from the published sequence for this
                transcript"
                /codon_start=1
                /db_xref="FLYBASE:FBan0002647"
                /db_xref="FLYBASE:FBgn0003068"
                /protein_id="AAF45804.1"
                /db_xref="GI:7290345"
                /translation="MEGGESTESTHNTKVSDSAYSNSCSNSQSQRSGSSKSRLSGSHS
                SGSSGYGGKPSTQASSSDMIIKRNKDKSRKKKKNKGAGQGAGQAQTLISASTSLEGRD
                EEKPRPSGTGCVEQQICRELQDQQHGEDHSEPQAIEQLQQEEEEDQSGSESEADRVEG
                VAKSEAAQSFPIPSPLSVTIVPPSMGGCGGVGHAAGLDSGLAKFDKTWEAGPGKLESM
                TGVGAAAAGTGQRGERVKEDSFCCVISMHDGIVLYTTPSITDVLGYPRDMWLGRSFID
                FVHLKDRATFASQITTGIPIAESRGSVPKDAKSTFCVMLRRYRGLKSGGFGVIGRPVS
                YEPFRLGLTFREAPEEARPDNYMVSNGTNMLLVICATPIKSSYKVPDEILSQKSPKFA
                IRHTATGIISHVDSAAVSALGYLPQDLIGRSIMDFYHHEDLSVMKETYETVMKKGQTA
                GASFCSKPYRFLIQNGCYVLLETEWTSFVNPWSRKLEFVVGHHRVFQGPKQCNVFEAA
                PTCKLKISEEAQSRNTRIKEDIVKRLAETVSRPSDTVKQEVSRRCQALASFMETLMDE
                VSRADLKLELPHENELTVSERDSVMLGEISPHHDYYDSKSSTETPPSYNQLNYNENLL
                RFFNSKPVTAPAELDPPKTEPPEPRGTCVSGASGPMSPVHEGSGGSGSSGNFTTASNI
                HMSSVTNTSIAGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGNGT
```

```
  Query Match           11.5%;  Score 45.4;  DB 3;  Length 300205;
  Best Local Similarity  51.8%;  Pred. No. 0.0088;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      49 aatagatgtgctaagatggtatgtttcggtgtgtattctttggttgcaagcaccaggcat 108
           || ||| |  ||   ||  |||||   |  || | ||  |||     |||||    | ||
Db 244916 AACAGAAGGCCTTTAATTATATGTGCAGTAGTTTGTTTGTTGAGCTTAAGCAAGCCGTAT 244857

Qy     109 gagttttggtgaatttaatcaagagagtaattcgttacaagaacttggggtttctgtgca 168
           | | |   |  |  |||| || |    |   |   | | |    ||| ||   |   |
Db 244856 AAATATAAATTCAGCTAATGAATAACAAATAGCCGAAAATGCTAATGGAGTAGATCAAAA 244797

Qy     169 acccaagaaaagtctgaaaataaccaaatttagaaaatgtaaggaatcaggggtttcaa 228
           | | || |||||  | |||  ||  | | ||| || | ||||  ||| | ||  | |  |
Db 244796 ATCAAAAAAAAGAATAAAACAAATTATAATTATAATAAGTAATAAATAAAGGAATATGTA 244737

Qy     229 gactctagatagcaagaag 247
           |  | |||| || |  ||
Db 244736 TATACCAGATGGCGACGAG 244718
```

```
RESULT  10
AC011029/c
LOCUS       AC011029              161123 bp    DNA     linear   PRI 03-JUL-2001
DEFINITION  Homo sapiens chromosome 11, clone RP11-202L1, complete sequence.
ACCESSION   AC011029
VERSION     AC011029.7  GI:14589620
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161123)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-202L1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161123)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 161123)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S.,
            Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:14547822.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2161
                Center clone name: 202_L_1
            --------------.
FEATURES                Location/Qualifiers
     source             1. .161123
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="11"
                        /map="11"
                        /clone="RP11-202L1"
                        /clone_lib="RPCI-11 Human Male BAC"
     repeat_region      1. .300
                        /rpt_family="MLT2B1"
     repeat_region      301. .429
```

```
                     /rpt_family="LTR41"
repeat_region        600. .687
                     /rpt_family="CT-rich"
repeat_region        925. .945
                     /rpt_family="AT_rich"
repeat_region        1167. .1197
                     /rpt_family="AT_rich"
repeat_region        complement(1362. .1400)
                     /rpt_family="L1MEd"
repeat_region        complement(1401. .1453)
                     /rpt_family="MLT2D"
repeat_region        1454. .1488
                     /rpt_family="(TG)n"
repeat_region        complement(1489. .1869)
                     /rpt_family="MLT2D"
repeat_region        complement(1870. .2028)
                     /rpt_family="L1MEd"
repeat_region        complement(2082. .2485)
                     /rpt_family="MLT2C1"
repeat_region        complement(2500. .2628)
                     /rpt_family="MLT2C1"
repeat_region        complement(2641. .2919)
                     /rpt_family="L1MEd"
repeat_region        2924. .2951
                     /rpt_family="(TG)n"
repeat_region        2952. .2987
                     /rpt_family="(GA)n"
repeat_region        complement(3482. .3621)
                     /rpt_family="L1MEd"
repeat_region        3623. .4372
                     /rpt_family="L2"
repeat_region        4373. .4651
                     /rpt_family="AluSc"
repeat_region        4652. .4777
                     /rpt_family="L2"
repeat_region        complement(4781. .4850)
                     /rpt_family="L1M4"
repeat_region        4970. .5003
                     /rpt_family="(CA)n"
repeat_region        complement(5146. .5369)
                     /rpt_family="L1M4"
repeat_region        5703. .5745
                     /rpt_family="(TTTTTG)n"
repeat_region        complement(5747. .11749)
                     /rpt_family="L1PA2"
repeat_region        11921. .11956
                     /rpt_family="AT_rich"
repeat_region        11999. .12030
                     /rpt_family="AT_rich"
repeat_region        complement(13210. .13258)
                     /rpt_family="MLT2B1"
repeat_region        13280. .13348
                     /rpt_family="(TA)n"
repeat_region        complement(13349. .13753)
                     /rpt_family="MLT2B1"
repeat_region        15927. .15956
                     /rpt_family="AT_rich"
repeat_region        16097. .16172
                     /rpt_family="CT-rich"
repeat_region        16192. .16348
                     /rpt_family="CT-rich"
repeat_region        16448. .16540
                     /rpt_family="L2"
repeat_region        16912. .17438
                     /rpt_family="MER1A"
repeat_region        17455. .17538
                     /rpt_family="MER5A"
repeat_region        complement(18371. .18434)
                     /rpt_family="MLT1E1"
repeat_region        18435. .18778
                     /rpt_family="L1MC3"
repeat_region        18816. .19296
                     /rpt_family="L1MC3"
repeat_region        19298. .19400
                     /rpt_family="MLT1F1"
repeat_region        19401. .19780
                     /rpt_family="L1MC3"
repeat_region        19782. .19846
                     /rpt_family="L1MC3"
repeat_region        20261. .20304
                     /rpt_family="AT_rich"
repeat_region        20617. .21186
                     /rpt_family="L1MC5"
repeat_region        22044. .22655
                     /rpt_family="L1MC5"
repeat_region        complement(22656. .22950)
                     /rpt_family="AluSx"
repeat_region        22951. .23798
                     /rpt_family="L1MC5"
repeat_region        complement(23854. .23985)
                     /rpt_family="MER94"
repeat_region        24021. .24649
                     /rpt_family="L1MC5"
unsure               25038. .25066
                     /note="qual <30"
repeat_region        26219. .26265
                     /rpt_family="AT_rich"
repeat_region        27280. .27357
                     /rpt_family="(CATATA)n"
repeat_region        27521. .27546
                     /rpt_family="(CA)n"
repeat_region        27615. .27643
                     /rpt_family="AT_rich"
repeat_region        27980. .28267
                     /rpt_family="AluSx"
repeat_region        complement(28770. .28923)
                     /rpt_family="MER41B"
repeat_region        28924. .28985
                     /rpt_family="(CA)n"
repeat_region        complement(28986. .29284)
                     /rpt_family="MER41B"
repeat_region        29472. .29637
                     /rpt_family="L2"
repeat_region        29802. .30094
                     /rpt_family="AluSx"
repeat_region        30464. .30601
                     /rpt_family="MER117"
repeat_region        30767. .30803
                     /rpt_family="AT_rich"

  Query Match           11.3%;  Score 44.6;  DB 9;  Length 161123;
  Best Local Similarity   57.6%;  Pred. No. 0.015;
  Matches   80;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     120 aatttaatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaa 179
           ||   ||   | |||  |||   ||| |||||          || | ||||  | ||||
Db   63462 AAACAAAAATAAAGAAAAATAATTTAAAAGAATAAACAAAGCCTATACAACATATGAAAC 63403

Qy     180 gtctgaaaataaccaaatttagaaaatgtaaggaatcagggggtttcaagactctagata 239
            ||  ||||| |||||||||  ||| | |  ||  ||   | || | |||||   | | |
Db   63402 ATCATAAAATTACCAAATTTTCAAATTTTTTGGTGTCCCAGAGTGTGAAGACAAAATAAA 63343

Qy     240 gcaagaagtggaaaatcct 258
           ||||  |  ||| ||||
Db   63342 AGAAGATTTAAAAATTCCT 63324


RESULT  11
AL355502
LOCUS       AL355502              150152 bp    DNA     linear   HTG 13-JUN-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-477B18, *** SEQUENCING IN
            PROGRESS ***, 23 unordered pieces.
ACCESSION   AL355502
VERSION     AL355502.6  GI:9931755
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150152)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 28, 2000 this sequence version replaced gi:9863722.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA477B18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 139434 bases at least Q40
            Consensus quality: 142362 bases at least Q30
            Consensus quality: 144491 bases at least Q20
            Insert size: 147952; sum-of-contigs
            Insert size: 175899; 0.8% error; agarose-fp
            Quality coverage: 3.55x in Q20 bases; sum-of-contigs Quality
            coverage: 3.37x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3553: contig of 3553 bp in length
            *     3554 3653: gap of      100 bp
            *     3654     9751: contig of 6098 bp in length
            *     9752 9851: gap of      100 bp
            *     9852    17450: contig of 7599 bp in length
            *    17451 17550: gap of      100 bp
            *    17551    22901: contig of 5351 bp in length
            *    22902 23001: gap of      100 bp
            *    23002    27484: contig of 4483 bp in length
            *    27485 27584: gap of      100 bp
            *    27585    30159: contig of 2575 bp in length
            *    30160 30259: gap of      100 bp
            *    30260    33956: contig of 3697 bp in length
            *    33957 34056: gap of      100 bp
            *    34057    37859: contig of 3803 bp in length
            *    37860 37959: gap of      100 bp
            *    37960    43058: contig of 5099 bp in length
            *    43059 43158: gap of      100 bp
            *    43159    50268: contig of 7110 bp in length
            *    50269 50368: gap of      100 bp
            *    50369    56506: contig of 6138 bp in length
            *    56507 56606: gap of      100 bp
            *    56607    69070: contig of 12464 bp in length
            *    69071 69170: gap of      100 bp
            *    69171    81994: contig of 12824 bp in length
            *    81995 82094: gap of      100 bp
            *    82095    85189: contig of 3095 bp in length
            *    85190 85289: gap of      100 bp
            *    85290    94643: contig of 9354 bp in length
            *    94644 94743: gap of      100 bp
            *    94744    97325: contig of 2582 bp in length
            *    97326 97425: gap of      100 bp
            *    97426   106940: contig of 9515 bp in length
            *   106941 107040: gap of      100 bp
            *   107041   116218: contig of 9178 bp in length
            *   116219 116318: gap of      100 bp
            *   116319   119546: contig of 3228 bp in length
            *   119547 119646: gap of      100 bp
            *   119647   122474: contig of 2828 bp in length
            *   122475 122574: gap of      100 bp
            *   122575   129655: contig of 7081 bp in length
            *   129656 129755: gap of      100 bp
            *   129756   138583: contig of 8828 bp in length
            *   138584 138683: gap of      100 bp
            *   138684   150152: contig of 11469 bp in length.
FEATURES             Location/Qualifiers
     source          1. .150152
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-477B18"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .3553
                     /note="assembly_fragment:01614
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    3654. .9751
                     /note="assembly_fragment:01503
                     fragment_chain:1"
     misc_feature    9852. .17450
                     /note="assembly_fragment:01244
                     fragment_chain:1"
     misc_feature    17551. .22901
                     /note="assembly_fragment:01044
                     fragment_chain:1"
     misc_feature    23002. .27484
                     /note="assembly_fragment:01629
                     fragment_chain:2"
     misc_feature    27585. .30159
                     /note="assembly_fragment:01631
                     fragment_chain:2"
     misc_feature    30260. .33956
                     /note="assembly_fragment:00502
                     fragment_chain:2"
     misc_feature    34057. .37859
                     /note="assembly_fragment:01432
                     fragment_chain:2"
     misc_feature    37960. .43058
                     /note="assembly_fragment:00405
                     fragment_chain:2"
     misc_feature    43159. .50268
                     /note="assembly_fragment:01016
                     fragment_chain:3"
     misc_feature    50369. .56506
                     /note="assembly_fragment:00860
                     fragment_chain:3"
     misc_feature    56607. .69070
                     /note="assembly_fragment:00437
                     fragment_chain:3"
     misc_feature    69171. .81994
                     /note="assembly_fragment:01040
                     fragment_chain:3"
     misc_feature    82095. .85189
                     /note="assembly_fragment:00091
                     fragment_chain:4"
     misc_feature    85290. .94643
                     /note="assembly_fragment:00369
                     fragment_chain:4"
     misc_feature    94744. .97325
                     /note="assembly_fragment:00207
                     fragment_chain:4"
     misc_feature    97426. .106940
                     /note="assembly_fragment:00844
                     fragment_chain:5"
     misc_feature    107041. .116218
                     /note="assembly_fragment:00840
                     fragment_chain:5"
     misc_feature    116319. .119546
```

```
                     /note="assembly_fragment:00452
                     fragment_chain:5"
     misc_feature    119647. .122474
                     /note="assembly_fragment:01555
                     fragment_chain:6"
     misc_feature    122575. .129655
                     /note="assembly_fragment:00103
                     fragment_chain:6"
     misc_feature    129756. .138583
                     /note="assembly_fragment:01041"
     misc_feature    138684. .150152
                     /note="assembly_fragment:01043"
BASE COUNT    42791 a  29791 c  30984 g  44377 t   2209 others
ORIGIN


  Query Match             11.2%;  Score 44.2;  DB 2;  Length 150152;
  Best Local Similarity   50.2%;  Pred. No. 0.02;
  Matches  164;  Conservative    0;  Mismatches  158;  Indels    5;  Gaps    2;

Qy      69 atgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaatttaatc 128
           |||| ||  || |  || ||||| || ||||| |       |  | ||   | || || |
Db   52686 ATGTGTCACTGAGAGTTTTTTGGGTGGAAGCAACCAAATCCAAATCTGACTAGTTCAAAC 52745

Qy     129 aagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtctgaaaa 188
           ||| |  | ||  |   |     | |    ||   |  ||   |||||    ||||  |
Db   52746 TAGAAAAGATTTTTTGGAAGTTTATGGACTGTTTCATTAAATTGAAGAAGCAACTGACCA 52805

Qy     189 taaccaaatttagaaaatgtaaggaatcaggggggtttcaagactctagatagcaagaagt 248
           | |    ||| ||||  ||| | | |  ||  | | || || ||||||||    ||  |
Db   52806 -ACCAGGCTTTGGAAAGGGTAGGAATTAGGGTAGCTCCAGGAATCTAGATAATGGGACTT 52864

Qy     249 ----ggaaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggttta 304
               || |     |||  | ||  |   || | |    | | ||   ||    |  |||
Db   52865 GTTGGGCAGTATTTGTTCCTAGATACTATCATGAAAATAAACTCGCTTCAGTATTTTTTC 52924

Qy     305 tatccaagtgccattctccccaacctccaggagtgagagtttctttagtatgtcttggat 364
           || |  ||| || | | | |||| | ||       |    | | | |  ||  ||||
Db   52925 AATTCTTGTGTCACTGTGCTCAACTTTCAAATTCCTGCTACTGTCTGGCCTGGTTTGGGC 52984

Qy     365 cctatgcccactttgtgaccagggcaa 391
           |  ||||||  ||  || | || | ||
Db   52985 CACATGCCCGTTTCTTGGCAAGAGAAA 53011


RESULT  12
AC025578
LOCUS       AC025578               113559 bp    DNA     linear   PRI 23-MAR-2001
DEFINITION  Homo sapiens 12 BAC RP11-340M11 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC025578
VERSION     AC025578.25  GI:13173484
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 113559)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
            Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
            Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
            Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B.,
            Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
            Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113559)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 113559)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 113559)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 1, 2001 this sequence version replaced gi:13162261.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
```

clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

---------- Summary Statistics ----------

Contig length:	113559
Phrap values in estimate:	110137
Average error rate (BCM-Phrap estimate):	7.10641e-06
Fraction of Phrap values less than 40 :	0.00837139
Number of consensus changing edits:	464
Number of N's in consensus :	0

---------- Consensus changing edits ----------

Position	Original+Context	Edited+Context
16488	catctctcta(n)caatgatgac	catctctcta(g)caatgatgac
32441	ccgaggcaac(n)tagagagacc	ccgaggcaac(a)tagagagacc
33412	gagacatgaa(n)aatgtgtcgg	gagacatgaa(g)aatgtgtcgg
46798	attgaactcc(n)atgcggtaag	attgaactcc(a)atgcggtaag
50131	atagagaatg(n)tgagttaatg	atagagaatg(t)tgagttaatg
50158	agggctacta(n)tttttcagat	agggctacta(c)tttttcagat
50357	ctaatataaa(n)ctggtcttat	ctaatataaa(a)ctggtcttat
52629	attctgatcc(c)cctccccaga	attctgatcc(t)cctccccaga
52654	tcctggaggg(g)ataacatcac	tcctggaggg(t)ataacatcac
53121	tatatattct(n)cattcttctt	tatatattct(a)cattcttctt
53628	atatatatat(n)nntatatata	atatatatat(g)actatatata
53629	tatatatatn(n)ntatatatan	tatatatatg(a)ctatatataca
53630	atatatatnn(n)tatatatana	atatatatga(c)tatatataca
53639	nntatatata(n)agtcctatac	actatatata(c)agtcctatac
55323	cccctccctg(n)nnccnnnnnt	cccctccctg(t)gtccatgtgt
55324	ccctccctgn(n)nccnnnnntt	ccctccctgt(g)tccatgtgtt
55325	cctccctgnn(n)ccnnnnnttn	cctccctgtg(t)ccatgtgttc
55328	ccctgnnncc(n)nnnnttnnnn	ccctgtgtcc(a)tgtgttctca
55329	cctgnnnccn(n)nnnttnnnnt	cctgtgtcca(t)gtgttctcat
55330	ctgnnnccnn(n)nnttnnnntt	ctgtgtccat(g)tgttctcatt
55331	tgnnnccnnn(n)nttnnnnttn	tgtgtccatg(t)gttctcattg
55332	gnnnccnnnn(n)ttnnnnttnt	gtgtccatgt(g)ttctcattgt
55335	nccnnnnntt(n)nnnttnttna	tccatgtgtt(c)tcattgttca
55336	ccnnnnnttn(n)nnttnttnaa	ccatgtgttc(t)cattgttcaa
55337	cnnnnnttnn(n)nttnttnaan	catgtgttct(c)attgttcaac
55338	nnnnnttnnn(n)ttnttnaann	atgtgttctc(a)ttgttcaact
55341	nnttnnnntt(n)ttnaannccc	tgttctcatt(g)ttcaactccc
55344	tnnnnttntt(n)aannccccnt	tctcattgtt(c)aactcccact
55347	nnttnttnaa(n)ncccnnttnn	cattgttcaa(c)tcccacttat
55348	nttnttnaan(n)cccnnttnnn	attgttcaac(t)cccacttatg
55352	ttnaannccc(n)nttnnnnnnn	ttcaactccc(a)cttatgagtg
55353	tnaanncccn(n)ttnnnnnnnn	tcaactccca(c)ttatgagtga
55356	annccnnntt(n)nnnnnnnnaa	actcccactt(a)tgagtgagaa
55357	nnccnnnttn(n)nnnnnnnaan	ctcccactta(t)gagtgagaat
55358	ncccnnttnn(n)nnnnnnaann	tcccacttat(g)agtgagaata
55359	cccnnttnnn(n)nnnnnaannn	cccacttatg(a)gtgagaatat
55360	ccnnttnnnn(n)nnnnaannng	ccacttatga(g)tgagaatatg
55361	cnnttnnnnn(n)nnnaannngg	cacttatgag(t)gagaatatgg
55362	nnttnnnnnn(n)nnaannnggg	acttatgagt(g)agaatatggg
55363	nttnnnnnnn(n)naannnnggg	cttatgagtg(a)gaatatgggg
55364	ttnnnnnnnn(n)aannnggggn	ttatgagtga(g)aatatggggt
55367	nnnnnnnnaa(n)nnggggnnct	tgagtgagaa(t)atggggtgct
55368	nnnnnnnaan(n)nggggnnctt	gagtgagaat(a)tggggtgctt
55369	nnnnnnaann(n)ggggnncttg	agtgagaata(t)ggggtgcttg
55374	naannnggg(n)ncttggtttt	gaatatgggg(t)gcttggtttt
55375	aannnggggn(n)cttggttttc	aatatggggt(g)cttggtttc
55619	agaatgattt(n)naanccttg	agaatgattt(a)taatcctttg
55620	gaatgatttn(n)aanccttgg	gaatgattta(t)aatccttgg
55623	tgatttnnaa(n)cctttgggta	tgatttataa(t)cctttgggta
68644	atcctttagg(n)nngcagctac	atcctttagg(c)tggcagctac
68645	tcctttaggn(n)ngcagctaca	tcctttaggc(t)ggcagctaca
68646	cctttaggnn(n)gcagctacaa	cctttaggct(g)gcagctacaa
68754	acatacattt(n)nttntctcnt	acatacattt(c)attctctcat
68755	catacatttn(n)ttntctcntn	catacatttc(a)ttctctcatc
68758	acatttnntt(n)tctcntnttt	acatttcatt(c)tctcatcttt
68763	tnnttntctc(n)tntttnttnn	tcattctctc(a)tctttgttca
68765	nttntctcnt(n)tttnttnnnt	attctctcat(c)tttgttcact
68769	tctcntnttt(n)ttnnnttttt	tctcatcttt(g)ttcacttttt
68772	cntnttnntt(n)nnttttttcn	catctttgtt(c)actttttct
68773	ntnttttnttn(n)nttttttcnn	atctttgttc(a)ctttttcta
68774	tnttttnttnn(n)ttttttcnnn	tctttgttca(c)tttttctac
68782	nnnttttttc(n)nnctngnacn	cacttttttc(t)acctggtaca
68783	nnttttttcn(n)nctngnacna	actttttttct(a)cctggtacaa
68784	nttttttcnn(n)ctngnacnac	ctttttttcta(c)ctggtacaac
68787	ttttcnnnct(n)gnacnacctc	ttttctacct(g)gtacaacctc
68789	ttcnnnctng(n)acnacctctt	ttctacctgg(t)acaacctctt
68792	nnnctngnac(n)acctcttcnn	tacctggtac(a)acctcttctg
68801	cnacctcttc(n)nnncccnttc	caacctcttc(t)gtcccacttc
68802	nacctcttcn(n)nnccnctttct	aacctctttct(g)tcccacttct
68803	acctcttcnn(n)nccnctttctn	acctctttctg(t)cccacttcta
68804	cctcttcnnn(n)ccnctttctnc	cctctttctgt(c)ccacttctac
68807	cttcnnnnnc(n)ctttctnccttt	cttctgtccc(a)ctttctacctt
68813	nnccnctttct(n)ccttncattn	tcccacttct(a)ccttccatta
68818	cttctnccttt(n)cattntttttc	cttctaccttt(c)cattattttc
68823	nccttncatt(n)ttttcnctttt	accttccatt(a)ttttcacttc
68829	cattntttttc(n)ctttcttttt	cattattttc(a)cttcctttta

Query Match 10.9%; Score 43; DB 9; Length 113559;
Best Local Similarity 51.5%; Pred. No. 0.046;
Matches 123; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

```
Qy      71 gtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaatttaatcaa 130
           ||||   |   |||| |||||||  | ||| |||| || |  | ||   | || | || |
Db   49228 GTTTACATTAATATTTTTTGGTTATATGCAACAGGGATCAACTCTGTCCACTTAAGTCCA 49287

Qy     131 gagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtctgaaaata 190
           | | | | | | |   |  |    || || ||   |||  ||| | ||  |   | | |
Db   49288 AAAAATGAATTGATTAGAAGAAAATGGATTACTCATCAAATCAAAAGAAAACCTGACA-A 49346

Qy     191 accaaatttagaaaatgtaaggaatcaggggttttcaagactctagatagcaagaagtgg 250
           |    ||||||||    |  |    |    |||||    ||| || || || |  |
Db   49347 CCAGGCTTTAGAAATGACAGAAACCAGGAAATTTTCAGTGATCTGGAGAGAAACACCTAC 49406

Qy     251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcc 309
           | |||  | |   ||||||  | |      |   | ||||  | | |   ||| | |||
Db   49407 AGAATGGTATAGTCAGGGAGACACTGAGATGATTATTCAGTCCTAACCATTTTGTCTCC 49465
```

RESULT 13
AL359180/c
LOCUS AL359180 159229 bp DNA linear PRI 26-JAN-2001
DEFINITION Human DNA sequence from clone RP11-320G21 on chromosome 13, complete sequence.
ACCESSION AL359180
VERSION AL359180.17 GI:12584420
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159229)
AUTHORS Tromans,A.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jan 28, 2001 this sequence version replaced gi:12329389.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr13
RP11-320G21 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-320G21 The true right end of clone RP11-94N9 is at 22847 in this sequence.

FEATURES Location/Qualifiers

source 1. .159229
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-320G21"
/clone_lib="RPCI-11.2"

repeat_region 332. .386
/note="L1MC2 repeat: matches 6269. .6325 of consensus"

repeat_region 422. .511
/note="L1M4 repeat: matches 5307. .5398 of consensus"

repeat_region 508. .688
/note="L1PA8 repeat: matches 4808. .4988 of consensus"

repeat_region 691. .1859
/note="L1PA8 repeat: matches 4990. .6155 of consensus"

repeat_region 2024. .2545
/note="L1ME3A repeat: matches 5574. .6136 of consensus"

repeat_region 3110. .3537
/note="L1MA10 repeat: matches 5894. .6320 of consensus"

repeat_region 3899. .3958
/note="5 copies 12 mer 78% conserved"

repeat_region 3902. .3979
/note="13 copies 6 mer cttttc 66% conserved"

repeat_region 5497. .5949
/note="MLT1D repeat: matches 1. .505 of consensus"

repeat_region 6147. .6182
/note="3 copies 12 mer 94% conserved"

repeat_region 6150. .6179
/note="5 copies 6 mer ttatta 100% conserved"

repeat_region 7335. .7442
/note="MIR repeat: matches 36. .146 of consensus"

repeat_region 7802. .8018
/note="MIR repeat: matches 6. .240 of consensus"

repeat_region 8048. .8107
/note="10 copies 6 mer gagtag 73% conserved"

repeat_region 8051. .8104
/note="18 copies 3 mer gag 72% conserved"

repeat_region 10577. .10688
/note="MIR repeat: matches 151. .262 of consensus"

repeat_region 11719. .11778
/note="5 copies 12 mer 85% conserved"

repeat_region 11720. .11777
/note="29 copies 2 mer tc 82% conserved"

repeat_region 11781. .11816
/note="3 copies 12 mer 94% conserved"

repeat_region 12127. .12228
/note="MIR repeat: matches 50. .156 of consensus"

repeat_region 12732. .13280
/note="L1M1 repeat: matches 4580. .5151 of consensus"

repeat_region 13378. .13405
/note="14 copies 2 mer aa 89% conserved"

repeat_region 13428. .13941
/note="MLT1D repeat: matches 6. .505 of consensus"

repeat_region 15176. .15270
/note="MIR repeat: matches 116. .212 of consensus"

repeat_region 15346. .15578
/note="MIR repeat: matches 11. .249 of consensus"

repeat_region 16660. .16876
/note="MIR repeat: matches 17. .250 of consensus"

repeat_region 17259. .17523
/note="MSTA repeat: matches 170. .425 of consensus"

repeat_region 17524. .17632
/note="AluY repeat: matches 1. .109 of consensus"

repeat_region 17633. .17719
/note="Alu repeat: matches 221. .307 of consensus"

repeat_region 17741. .17887
/note="MSTA repeat: matches 18. .185 of consensus"

repeat_region 18109. .18394
/note="L2 repeat: matches 2423. .2750 of consensus"

repeat_region 18680. .18772
/note="L1M1 repeat: matches 5333. .5426 of consensus"

repeat_region 19170. .19393
/note="LTR33 repeat: matches 261. .521 of consensus"

repeat_region 19579. .19733
/note="MIR repeat: matches 20. .187 of consensus"

repeat_region 20080. .20389
/note="AluY repeat: matches 1. .310 of consensus"

repeat_region 20597. .21622
/note="L1MA8 repeat: matches 5216. .6271 of consensus"

repeat_region 21678. .22712
/note="L1ME repeat: matches 4501. .5583 of consensus"

repeat_region 22706. .22912
/note="L1PBa repeat: matches -1549. .-1343 of consensus"

repeat_region 22915. .23407
/note="L1PB1 repeat: matches -940. .-411 of consensus"

repeat_region 23403. .24054
/note="L1PB1 repeat: matches -265. .382 of consensus"

repeat_region 24055. .24774
/note="LTR8 repeat: matches 1. .691 of consensus"

repeat_region 24775. .24901
/note="L1PB1 repeat: matches 382. .1464 of consensus"

repeat_region 24904. .25150
/note="AluY repeat: matches 1. .299 of consensus"

repeat_region 25158. .27217
/note="L1PB1 repeat: matches 1456. .3548 of consensus"

repeat_region 27218. .27534
/note="AluY repeat: matches 1. .310 of consensus"

repeat_region 27535. .28414
/note="L1PB1 repeat: matches 3548. .4415 of consensus"

repeat_region 28426. .28701
/note="L1 repeat: matches 2235. .2511 of consensus"

repeat_region 28697. .30084
/note="L1PB1 repeat: matches 4416. .5790 of consensus"

repeat_region 30088. .30147
/note="30 copies 2 mer at 76% conserved"

repeat_region 30152. .30433
/note="L1PB1 repeat: matches 5781. .6070 of consensus"

repeat_region 30433. .30660
/note="L1ME3 repeat: matches 5849. .6075 of consensus"

repeat_region 30877. .31035
/note="MLT1A1 repeat: matches 1. .158 of consensus"

repeat_region 31036. .31286
/note="AluSg1 repeat: matches 1. .289 of consensus"

repeat_region 31287. .31490
/note="MLT1A1 repeat: matches 158. .365 of consensus"

repeat_region 31568. .31864
/note="AluY repeat: matches 1. .295 of consensus"

repeat_region 32106. .32189
/note="7 copies 12 mer 70% conserved"

repeat_region 32242. .32440
/note="MER6B repeat: matches 3. .210 of consensus"

repeat_region 32453. .32512
/note="30 copies 2 mer ta 70% conserved"

repeat_region 32858. .32881

```
                     /note="12 copies 2 mer ac 95% conserved"
     repeat_region   33008. .33372
                     /note="MLT1A1 repeat: matches 1. .361 of consensus"
     repeat_region   33786. .33842
                     /note="L2 repeat: matches 2688. .2744 of consensus"
     repeat_region   33788. .33856
                     /note="MIR repeat: matches 193. .259 of consensus"
     repeat_region   35886. .35919
                     /note="17 copies 2 mer aa 82% conserved"
     repeat_region   36181. .36480
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   38591. .38839
                     /note="L1M4 repeat: matches 2258. .2507 of consensus"
     repeat_region   38840. .39118
                     /note="AluSc repeat: matches 1. .284 of consensus"
     repeat_region   39119. .39182
                     /note="L1M4 repeat: matches 2507. .2570 of consensus"
     repeat_region   39199. .39874
                     /note="L1MA9 repeat: matches 5266. .5928 of consensus"
     repeat_region   39878. .39925
                     /note="4 copies 12 mer 83% conserved"
     repeat_region   39995. .40281
                     /note="AluJo repeat: matches 1. .284 of consensus"
     repeat_region   40282. .41092
                     /note="L1MA9 repeat: matches 4495. .5276 of consensus"
     repeat_region   41094. .41568
                     /note="L1 repeat: matches 3294. .3774 of consensus"
     repeat_region   41583. .41813
                     /note="L1 repeat: matches 2911. .3139 of consensus"
     repeat_region   41887. .42845
                     /note="L1 repeat: matches 1940. .2905 of consensus"
     repeat_region   42850. .43085
                     /note="L1MC/D repeat: matches 5277. .5507 of consensus"

  Query Match             10.8%;  Score 42.8;  DB 9;  Length 159229;
  Best Local Similarity   60.2%;  Pred. No. 0.053;
  Matches   71;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      70 tgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaatttaatca 129
           | | |  || |  |||||||||| || ||||| |||  ||| ||| |  | ||||||| ||
Db   36599 TTTCTGTGTCTAGATTCTTTGGCTGTAAGCAACAGAGATGGGTTCTAATTAATTTAAACA 36540

Qy     130 agagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtctgaaa 187
           ||| || | ||  ||   || |  | |  |  ||  |  | || | | ||     |||
Db   36539 AGACAGGATTTTATTGGGAGGAAATAGAATGGCTAAGGGATCCTAAAGAAAGGCTAAA 36482


RESULT  14
AL137009
LOCUS       AL137009               113254 bp    DNA     linear   PRI 22-NOV-2001
DEFINITION  Human DNA sequence from clone RP3-370O5 on chromosome 6, complete
            sequence.
ACCESSION   AL137009
VERSION     AL137009.8  GI:9581553
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 113254)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 28, 2000 this sequence version replaced gi:9407728.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP3-370O5 is from the
            library RPCI-3 constructed by the group of Pieter de Jong. For
            further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP3-370O5 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP3-370O5 is at 1 in this sequence. The
            true left end of clone RP3-354N19 is at 113155 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .113254
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-370O5"
                     /clone_lib="RPCI-3"
     repeat_region   59. .142
                     /note="42 copies 2 mer aa 63% conserved"
     repeat_region   13347. .13438
                     /note="HY4 repeat: matches 1. .93 of consensus"
     repeat_region   14409. .14462
                     /note="27 copies 2 mer tt 75% conserved"
     repeat_region   17232. .17290
                     /note="L1MA5A repeat: matches 5981. .6038 of consensus"
     repeat_region   17291. .17382
                     /note="L1MA5A repeat: matches 6199. .6294 of consensus"
     repeat_region   17385. .17497
                     /note="AluJ/monomer repeat: matches 1. .49 of consensus"
     repeat_region   22889. .22973
                     /note="MER82 repeat: matches 570. .653 of consensus"
     repeat_region   22978. .23169
                     /note="MER82 repeat: matches 83. .277 of consensus"
     misc_feature    33394. .34525
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   36771. .37360
                     /note="LTR1 repeat: matches 203. .785 of consensus"
     repeat_region   37393. .37617
                     /note="LTR1 repeat: matches 32. .258 of consensus"
     repeat_region   37769. .39375
                     /note="THE1C-internal repeat: matches 3. .1580 of
                     consensus"
     repeat_region   39378. .39755
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   39760. .41138
                     /note="MSTA-internal repeat: matches 1. .1442 of consensus"
     repeat_region   46143. .46588
                     /note="HERVL repeat: matches 4933. .5369 of consensus"
     repeat_region   46867. .47592
                     /note="HERVL repeat: matches 3966. .4698 of consensus"
     repeat_region   48033. .48515
                     /note="HERVL repeat: matches 1521. .2005 of consensus"
     repeat_region   48756. .48901
                     /note="HERV16 repeat: matches 949. .1123 of consensus"
     repeat_region   49431. .49550
                     /note="60 copies 2 mer ct 60% conserved"
     repeat_region   53032. .53393
                     /note="MER39 repeat: matches 294. .688 of consensus"
```

```
     repeat_region   53394. .53479
                     /note="LTR39 repeat: matches 711. .794 of consensus"
     repeat_region   53837. .54314
                     /note="LTR39 repeat: matches 217. .711 of consensus"
     repeat_region   54616. .54742
                     /note="MER21B repeat: matches 295. .450 of consensus"
     repeat_region   54742. .54964
                     /note="LTR39 repeat: matches 1. .226 of consensus"
     repeat_region   63591. .63618
                     /note="14 copies 2 mer tg 92% conserved"
     repeat_region   63766. .63813
                     /note="L1PA15-16 repeat: matches -209. .-158 of consensus"
     repeat_region   64934. .64969
                     /note="18 copies 2 mer tc 91% conserved"
     repeat_region   71439. .71486
                     /note="24 copies 2 mer ca 97% conserved"
     repeat_region   73918. .75001
                     /note="MER11C repeat: matches 1. .1071 of consensus"
     repeat_region   77802. .78110
                     /note="MER72 repeat: matches 382. .668 of consensus"
     repeat_region   78422. .78555
                     /note="MER72 repeat: matches 244. .382 of consensus"
     repeat_region   79066. .79195
                     /note="MER49 repeat: matches 2. .131 of consensus"
     repeat_region   81935. .82045
                     /note="LTR28 repeat: matches 848. .957 of consensus"
     repeat_region   82035. .82688
                     /note="LTR1 repeat: matches 1. .662 of consensus"
     repeat_region   82689. .82782
                     /note="MLT2E repeat: matches 44. .137 of consensus"
     repeat_region   83634. .84159
                     /note="LTR38 repeat: matches 1. .547 of consensus"
     repeat_region   92833. .92932
                     /note="MLT1A2 repeat: matches 68. .170 of consensus"
     repeat_region   92937. .93048
                     /note="MLT1A2 repeat: matches 1. .115 of consensus"
     repeat_region   100454. .100649
                     /note="LTR33 repeat: matches 1. .219 of consensus"
     repeat_region   101865. .101906
                     /note="21 copies 2 mer gt 100% conserved"
     repeat_region   102970. .103521
                     /note="MER89 repeat: matches 1. .557 of consensus"
     repeat_region   111796. .111962
                     /note="L1PA5 repeat: matches 5979. .6145 of consensus"
BASE COUNT    33925 a  23518 c  23177 g  32634 t
ORIGIN


  Query Match           10.8%;  Score 42.6;  DB 9;  Length 113254;
  Best Local Similarity   49.8%;  Pred. No. 0.061;
  Matches  163;  Conservative    0;  Mismatches  159;  Indels    5;  Gaps    2;

Qy      69 atgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaatttaatc 128
           |||| ||  || |  || ||||| || ||||| |       |  | ||   | || || |
Db   91957 ATGTGTCACTGAGAGTTTTTTGGGTGGAAGCAACCAAATCCAAATCTGACTAGTTCAAAC 92016

Qy     129 aagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtctgaaaa 188
            ||| |  | ||  |   |     | |    ||   |  ||   |||||    |||| |
Db   92017 TAGAAAAGATTTTTTGGAAGTTTATGGACTGTTTCATTAAATTGAAGAAGCAACTGACCA 92076

Qy     189 taaccaaatttagaaaatgtaaggaatcaggggttttcaagactctagatagcaagaagt 248
            | |    ||| ||||  ||| | | |  ||  | | || || ||||||||    ||  |
Db   92077 -ACCAGGCTTTGGAAAGGGTAGGAATTAGGGTAGCTCCAGGAATCTAGATAATGGGACTT 92135

Qy     249 ----ggaaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggttta 304
               || |     |||  | ||  |   || | |    | | ||   ||    |  |||
Db   92136 GTCGGGCAGTATTTGTTCCTAGATACTATCATGAAAATAAACTCGCTTCAGTATTTTTTC 92195

Qy     305 tatccaagtgccattctccccaacctccaggagtgagagtttctttagtatgtcttggat 364
            || |  ||| || | | |  ||| | ||       |    | | | |  ||  ||||
Db   92196 AATTCTTGTGTCACTGTGCTTAACTTTCAAATTCCTGCTACTGTCTGGCCTGGTTTGGGC 92255

Qy     365 cctatgcccactttgtgaccagggcaa 391
```

```
           |  ||||||  ||  || | || | ||
Db   92256 CACATGCCCGTTTCTTGGCAAGAGAAA 92282


RESULT  15
HUAC002302
LOCUS       HUAC002302            259894 bp    DNA     linear   PRI 23-NOV-1999
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987-SKA-345G4  ~complete
            genomic sequence, complete sequence.
ACCESSION   AC002302
VERSION     AC002302.1  GI:2576341
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 259894)
  AUTHORS   Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
            Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
            Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
            Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
  TITLE     Genome duplications and other features in 12 Mb of DNA sequence
            from human chromosome 16p and 16q
  JOURNAL   Genomics 60 (3), 295-308 (1999)
  MEDLINE   99425270
   PUBMED   10493829
REFERENCE   2  (bases 1 to 259894)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., LaBombard,M., Fuhrmann,J.,
            Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987-SKA-345G4 #complete
            genomic sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 259894)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-1997) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA, Email:
            mdadams@tigr.org
REFERENCE   4  (bases 1 to 259894)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-1997) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   5  (bases 1 to 259894)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1997) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Oct 30, 1997 this sequence version replaced gi:2570249.
            Address all correspondence to:
            Mark Adams
            The Institute for Genomic Research
            9712 Medical Center Dr,
            Rockville, MD 20850,
            USA
            e-mail address: mdadams@tigr.org.  The bac location is on
            chromosome BAC clone  is located on human chromosome 16p12.2-p12 .

            The orientation of the sequence is from SP6 end to T7 end.
            Genes were identified by a combination of five methods including:
            XGRAIL (available by anonymous ftp from arthur.epm.ornl.gov),
            Genefinder (Phil Green, University of Washington), Genscan (Chris
            Burge, http://gnomic.stanford.Edu/~chris/GENSCANW.html
            )
            searches of the complete sequence against a peptide database,
            and the Human gene Index database at TIGR
            (http://www.tigr.org/tdb/hg
            i/hgi.html).
            A gene with homolgy to another protein is annotated as the isolog
            of
            that protein.
            Genes without pepetide homolgy having spliced EST hits are termed
            'u
```

```
        nknown protein'.
        Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
        http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1. .259894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p12.2-p12"
                     /clone="A-345G4"
     repeat_region   complement(687. .988)
                     /rpt_family="AluSp"
     STS             782. .899
                     /db_xref="dbSTS:G02122"
     repeat_region   complement(989. .1290)
                     /rpt_family="AluJo"
     mRNA            join(2534. .2600,3145. .3217,3331. .3411,3547. .3677,
                     3872. .3933,4686. .4808,5015. .5206)
                     /gene="A-345G4"
     gene            2534. .5206
                     /gene="A-345G4"
     CDS             join(2534. .2600,3145. .3217,3331. .3411,3547. .3677,
                     3872. .3933,4686. .4808,5015. .5068)
                     /gene="A-345G4"
                     /codon_start=1
                     /product="Unknown protein product with similarity to
                     Calcineurin B subunit"
                     /protein_id="AAB91429.1"
                     /db_xref="GI:2695572"
                     /translation="MGSRSSHAAVIPDGDSIRRETGFSQASLLRLHHRFRALDRNKKG
                     YLSRMDLQQIGALAVNPLGDRIIESFFPDGSQRVDFPGFVRVLAHFRPVEDEDTETQD
                     PKKPEPLNSRRNKLHYAFQLYDLDRDGKISRHEMLQVLRLMVGVQVTEEQLENIADRT
                     VQEADEDGDGAVSFVEFTKSLEKMDVEQKMSIRILK"
     repeat_region   4139. .4358
                     /rpt_family="MER20"
     repeat_region   complement(5235. .5705)
                     /rpt_family="MER1A"
     repeat_region   complement(6635. .6810)
                     /rpt_family="MER63A"
     repeat_region   complement(6847. .6926)
                     /rpt_family="(GAAA)n"
     repeat_region   complement(7198. .7495)
                     /rpt_family="AluJb"
     repeat_region   7784. .8074
                     /rpt_family="AluSx"
     repeat_region   8081. .8116
                     /rpt_family="(CAA)n"
     repeat_region   complement(8463. .8627)
                     /rpt_family="AluJo/FRAM"
     repeat_region   complement(8630. .8800)
                     /rpt_family="AluSg/x"
     repeat_region   complement(8801. .9102)
                     /rpt_family="AluSx"
     repeat_region   complement(9103. .9165)
                     /rpt_family="AluJ/FLAM"
     repeat_region   9822. .9854
                     /rpt_family="AT_rich"
     repeat_region   9976. .10072
                     /rpt_family="L1MA5"
     repeat_region   10074. .10366
                     /rpt_family="AluSq"
     repeat_region   complement(10541. .10756)
                     /rpt_family="AluJo"
     repeat_region   11513. .11855
                     /rpt_family="MER7A"
     repeat_region   complement(11984. .12277)
                     /rpt_family="AluSx"
     repeat_region   12403. .12531
                     /rpt_family="MIR"
     repeat_region   complement(12806. .13105)
                     /rpt_family="AluSx"
     repeat_region   complement(13176. .13463)
                     /rpt_family="L1PA13"
     STS             13594. .13721
                     /db_xref="dbSTS:Z16998"
     repeat_region   complement(13624. .13660)
                     /rpt_family="(CA)n"
     repeat_region   14128. .14399
                     /rpt_family="AluSx"
     repeat_region   14814. .15051
                     /rpt_family="AluJo"
     repeat_region   15054. .15354
                     /rpt_family="AluY"
     repeat_region   15372. .15647
                     /rpt_family="AluJb"
     repeat_region   15740. .15760
                     /rpt_family="AT_rich"
     repeat_region   15891. .15925
                     /rpt_family="AT_rich"
     repeat_region   15929. .16230
                     /rpt_family="AluSx"
     repeat_region   16528. .16609
                     /rpt_family="LTR16A"
     repeat_region   complement(16655. .16889)
                     /rpt_family="AluJb"
     repeat_region   16911. .17120
                     /rpt_family="LTR16B"
     repeat_region   complement(17675. .17995)
                     /rpt_family="AluJo"
     repeat_region   18089. .18162
                     /rpt_family="(TAAAA)n"
     repeat_region   complement(18350. .18611)
                     /rpt_family="AluJo"
     repeat_region   18632. .18852
                     /rpt_family="AluSq"
     repeat_region   complement(19349. .19709)
                     /rpt_family="MER54"
     repeat_region   complement(19713. .20003)
                     /rpt_family="AluSq"
     repeat_region   complement(20006. .20080)
                     /rpt_family="MER54"
     repeat_region   20189. .20317
                     /rpt_family="(TA)n"
     repeat_region   complement(20354. .20480)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(20532. .20654)
                     /rpt_family="MIR"
     repeat_region   complement(20964. .21133)
                     /rpt_family="LINE2"
     repeat_region   21137. .21425
                     /rpt_family="AluJo"
     repeat_region   complement(21453. .21711)
                     /rpt_family="LINE2"
     repeat_region   complement(22022. .22309)
                     /rpt_family="AluSx"
     repeat_region   complement(22924. .23199)
                     /rpt_family="AluSg"
     repeat_region   complement(23202. .23228)
                     /rpt_family="AT_rich"
     repeat_region   complement(23719. .24017)
                     /rpt_family="AluJo"
     repeat_region   24067. .24363
                     /rpt_family="MLT1F"

  Query Match           10.7%;  Score 42.4;  DB 9;  Length 259894;
  Best Local Similarity  52.6%;  Pred. No. 0.071;
  Matches  120;  Conservative    0;  Mismatches  101;  Indels    7;  Gaps    1;

Qy      69 atgtttcggtgtgtattctttggttgcaagcaccaggcatgagttttggtgaatttaatc 128
           ||  ||  ||  | || |||||||||||||| | ||   ||||  | || | ||| ||| |
Db 124650 ATCCTTTAGTTAGAATCCTTTGGTTGCAAGTAACATAAATGAACTCTGATTAATGTAAGC 124709

Qy     129 aagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaagtctgaaaa 188
           || | || | ||  |   | ||  | ||||     |      |  |||   |   |||||
Db 124710 AAAAAAGAATTTGTTAGAAGGATATGGGGGGGGAGGGATGGACAGAGATGTGAAGGAAAA 124769
```

```
Qy     189 taaccaaa-------tttagaaaatgtaaggaatcaggggtttcaagactctagatagc 241
           |   | |       |||||||| |       |   |  |  ||  | |  |||||  |||
Db  124770 GATGAAGAACCAGGCTTTAGAAAGTCCTGAAACCAAATGAATTCTAGGGATCTAGGAAGC 124829

Qy     242 aagaagtggaaaatcctgtctccagggaacctcctcaggggagaatca 289
           | | | |    | |  |  || ||||  |||      |||| |||| |
Db  124830 AGGGACTAATGACTGTTTCCTTCAGGACACCAGAGATGGGGTGAATAA 124877


Search completed: August 26, 2002, 23:11:39
Job time: 6339 sec
```